

Lingvistisk skriftteori og bokmål

Bård Uri Jensen



Institutt for lingvistiske og nordiske studier

UNIVERSITETET I OSLO

22. april 2005

Skrift brukes hele tiden til å representere språk.
Vi kan ikke bare se bort fra den. Vi må være kjent
med dens nytte, mangler og farer.

Ferdinand de Saussure

Innhold

Innhold.....	3
Forord.....	5
1 Om oppgaven.....	6
1.1 Bakgrunn og formål.....	6
1.2 Notasjon.....	7
1.3 Metode og datagrunnlag.....	11
1.4 Resten av oppgaven.....	21
2 Avgrensninger.....	23
2.1 Standard østnorsk.....	23
2.2 Bokmål.....	25
2.3 Allomorfi.....	36
2.4 Synkronisitet.....	36
2.5 En lingvistisk undersøkelse.....	37
3 Strukturen i standard østnorsk.....	40
3.1 Fonologi.....	40
3.2 Morfologi og morfofonemikk.....	45
4 Grunnleggende skrifteori.....	47
4.1 Skrift og språk.....	48
4.2 Grafemet.....	55
4.3 Skriftypologi.....	65
4.4 Grafemmengde og alfabet.....	72
4.5 Fonemografi.....	83
5 Bokmål som fonemografisk skrift.....	116
5.1 Vokaler.....	116
5.2 Diftonger.....	118
5.3 Konsonanter.....	121
5.4 Konsonantsekvenser og digrafer.....	129
5.5 Allofoni	132
5.6 Prosodemer.....	132

5.7	Stavelsen.....	134
5.8	Intonasjon.....	137
6	Andre perspektiver på bokmål.....	141
6.1	Syllabografi.....	141
6.2	Morfemografi.....	144
6.3	Et generativt perspektiv.....	149
7	Kvantitative undersøkelser.....	154
7.1	Grunnleggende diskusjoner.....	154
7.2	Data.....	155
7.3	Metode.....	158
7.4	Resultater.....	161
7.5	Diskusjon.....	167
	Litteratur.....	169
	Importord.....	173
	Register.....	175

Forord

Dette arbeidet er levert som obligatorisk skriftlig arbeid – hovedoppgave – til graden cand.philol. ved Universitetet i Oslo våren 2005.

Takk til Kjell Ivar Vannebo, som har vært veileder for denne hovedoppgaven, Geirr Wiggen og Arne Torp, som var nyttige samtalepartnere i startfasen av prosjektet, Chr.-Emil Ore ved Dokumentasjonsprosjektet, som har gitt meg tilgang til datafilene bak Bokmålsordboka, Lars Nygård og Anders Nøklestad ved Tekstlaboratoriet, som har gitt meg tilgang til Oslo-korpuset, Jan-Olav Eide, som har svart på spørsmål av matematisk og IT-faglig karakter, Pål Erik Jensen, som, kanskje uten å vite det, har bidratt med diskusjoner og synspunkter, Lars Anders Kulbrandstad og Marianne Haslev Skånland, som har hatt tro på meg gjennom mange års studier. Og takk til familien min: Elin, Endre og Signhild, som har utvist tålmodighet med en middelaldrende mann som hadde lyst til å ta hovedfag.

Denne oppgaven er i hovedsak skrevet med det gratis tilgjengelige tekst-behandlingsprogrammet Writer fra OpenOffice.org og dets innebygde bibliografihåndteringsmodul. Dataene er hovedsakelig behandlet i regneark-programmet Calc fra OpenOffice.org og med et program av `grep`-familien.

1 Om oppgaven

1.1 Bakgrunn og formål

Vi lever i et land med stor oppmerksomhet rundt rettskrivning. Vi har rettskrivningsnormer som er *vedtatt* av offentlige instanser, ja landet vårt har endog et system av to parallelle rettskrivningsnormer for ett og samme språk. Vi har en lov som regulerer offentlige ansattes bruk av disse normene. Vi har frivillige organisasjoner som engasjerer seg i normenes utforming og bruk, og vi har frivillige organisasjoner som utformer sine egne, private rettskrivningsnormer. I skolen lærer elevene at nyere språkhistorie stort sett har med utvikling av rettskrivningsnormer å gjøre, og med korte mellomrom setter et utvalg av landets kloke hoder seg sammen for å justere disse normene.

I lys av all denne oppmerksomheten er det oppsiktsvekkende hvor lite synkront vinklet forskning som finnes på lingvistisk skrifteori og på rettskrivningsnormene som system. Mye oppmerksomhet er viet de nåværende rettskrivningssystemenes historie, og det er også forsket på barns rettskrivningsavvik og andre aspekter ved lesing og skriving, særlig hos barn. Ellers dreier mye av forskningen på skrift seg om tekstlingvistiske problemstillinger og disse problemstillingenes pedagogiske og didaktiske konsekvenser.

Denne hovedoppgaven etablerer og diskuterer grunnleggende skrifteoretiske begreper og gir en synkron, lingvistisk undersøkelse av bokmålets ortografiske system og dette systemets relasjoner til talespråkets system på det fonologiske, morfologiske og leksikalske nivå, på linje med hva som er gjort for engelsk av Venezky 1970 og til dels Chomsky og Halle 1968, og for svensk av Allén 1971. Arbeidet baserer seg for øvrig på generelle skrifteoretiske arbeider av Gelb 1963, Sampson 1985, DeFrancis 1989, Harris 1995, Sproat 2000 og Coulmas 2003.

Formålet med arbeidet er således å øke kunnskapen om skrift og om bokmålets lingvistiske egenskaper, både som et internt system og som en representasjon av

det talte språket. Det er min oppfatning at vitenskapelig og praktisk arbeid innen områder som normering, leseopplæring, andrespråksopplæring og lese- og skrivevansker vil ha fordel av mer omfattende og mer systematisk kunnskap om skriftens lingvistiske egenskaper generelt og bokmålets lingvistiske egenskaper spesielt, og oppgaven søker å fylle dette behovet.

Saussure [1915] fremhever dessuten skriftens rolle i studiet av språk: "Skrift [...] brukes hele tiden til å representere språk. Vi kan ikke bare se bort fra den. Vi må være kjent med dens nytte, mangler og farer."¹ Mer systematisk kunnskap om forholdet mellom skrift og tale vil kunne være et verktøy innenfor lingvistikken, også om man ikke primært studerer skriftrelaterte emner.

1.2 Notasjon

1.2.1 Transskripsjonsklammer

/.../ fonemisk transskripsjon. Transskripsjonen benytter stort sett IPA, men med noen tilpasninger. Se 1.2.2.

[...] fonetisk transskripsjon. Dette kan være en mer finmasket notasjon enn den fonemiske, men ofte vil det være en mer grovmasket notasjon, alt etter formålet med transskripsjonen.

"..." "semantisk transskripsjon." Et uttrykk for innholdssida i språktegnet.

<...> grafisk "transskripsjon". Klammene markerer at vi snakker om det skriftlige uttrykket. Denne gjengivelsen vil hovedsakelig være grafemisk, men i enkelte tilfeller vil uttrykket ikke ha grafemisk status, men fokusere på andre aspekter ved skriften. Dette vil gå fram av sammenhengen.

{...} har flere bruksområder, blant annet markering av dypfonologiske enheter og morfemer. Klammene er brukt der det er nødvendig å skille visse typer av enheter fra enhetene over med egne transskripsjonsklammer. Det er

¹ Min oversettelse fra den engelske utgaven.

begrenset hvor mange ulike klammetyper det er formålstjenlig å bruke i én og samme tekst, og det vil gå fram av sammenhengen hva slags type enheter som står i disse klammene.

kursiv Ved omtale av leksemer eller ordformer der det er unaturlig å fokusere på enten innholdet eller ett av uttrykkene, blir leksemene eller ordformene gjengitt i kursiv.

| _ kontekstmarkering. Både fonologisk og grafonomisk kontekst kan spesifiseres på denne måten, for eksempel: <t>|o_t, som vil representere ord som <flott>, <grotte>. Dersom det ikke er spesifisert, vil konteksten gjelde det samme uttrykksnivået som uttrykket foran kontekstmerket, eller uttrykksnivået vil være irrelevant. Det er imidlertid også mulig å markere kontekst fra det andre uttrykksnivået, for eksempel: <t>|[å]_, som vil representere ord som <gåte> og <rotte>.

¬ blokkering. Markerer at en viss enhet eller regel ikke kan forekomme.

ordgrense. Ordgrense kan markeres for både fonologiske og grafonomiske omgivelser. Symbolet kan brukes i kontekstmarkering f.eks. slik at <v>|#h_ representerer ord som <hval>, <hva> og <hvorfor>, men ikke <blåhval>.

Der det ikke er fare for misforståelser, forenkles formler som <h>|#_v til #<hv>.

+ junktur. Sammensetning, avledning eller bøyning. F.eks. *bre+is*, *klatr+ing* eller *is+en*.

- stavelsesgrense i fonetisk eller fonemisk transkripsjon. F.eks. /²u-tøy/ eller /²ut-øy/.

Bindestrek kan også brukes for å markere hele stavelser i fonetiske transskripsjon. F.eks. <rottene>=[²---], <ballett>=[-¹let].

K en hvilken som helst konsonant. F.eks <bil>=[KiK].

- V en hvilken som helst vokal. F.eks. <bil>=[bVl].
- markerer kort vokal ved behov. F.eks. <o>=/å'/_<tt>. Dette er ikke i tråd verken med IPA eller tradisjonen, men det viste seg vanskelig å kombinere transskripsjonssymbolet 'å' med et tradisjonelt symbol for kort vokal.

1.2.2 Fonemisk transskripsjon

En del av den transskripsjonen som blir brukt i oppgaven, blir gjort til gjenstand for diskusjon som vil problematisere bruken av de ulike symbolene. Således vil for eksempel [ɾ] og [l] delvis opptre som selvstendige fonemer og delvis som allofoner av /l/.

Oppgaven bruker i stor grad standardsymboler for fonemene i standard østnorsk, men med noen vanlige tilpasninger som vil være kjent for leseren og utvetydige i sammenhengen. Imidlertid vil jeg særlig påpeke at jeg representerer de tre ikke-fremre rundede vokalene med symbolene /u/, /o/, /å/, i stedet for /ʊ/, /ʊ/, /o/, som ville ha vært mer i tråd med IPA. Jeg er prinsipielt tilhenger av å følge IPA-standardene og mener generelt at dette skaper færrest misforståelser og færrest problemer eller heftelser for leseren. Imidlertid har jeg funnet at når symbolene for fonemene i så stor grad som i denne oppgaven opptre i kompaniskap med omtale av *grafemene* <u>, <o>, <å>, blir det lettest å lese og sammenhengene lettest å se med den løsningen jeg har valgt.

Jeg bruker også /ʃ/ i stedet for /s/, som kanskje er fonetisk mer korrekt.

Imidlertid benytter oppgaven nesten utelukkende fonemisk transskripsjon, eller en vid fonetiske transskripsjon der den artikulatoriske realiseringen ikke er vesentlig, og jeg velger da å benytte det symbolet som jeg oppfatter som å være vanligst i bruk. I fonemikken står som kjent funksjonen og ikke de fysiske egenskaper i sentrum.

Til slutt blant segmentalene vil jeg nevne diftongene, som jeg transskriberer med symbolkombinasjoner som kanskje er uvante. Jeg indikerer diftongen under ved hjelp av den mest vanlige stavemåten for diftongen:

- <ai>= /aj/ som i <hai>=/haj/
- <au>= /æu/ som i <haug>=/hæu/
- <ei>= /æj/ som i <hei>=/hæj/
- <ui>= /uj/ som i <huie>=/²huje/
- <oi>= /oj/ som i <hoie>=/²hoje/
- <oi>= /åj/ som i <koie>=/²kåje/
- <øy>= /øj/ som i <høye>=/²høje/

Kvantitet markeres i tråd med tradisjonen som kolon etter lang vokal. Tonem noteres som superskriberte tall foran den trykksterke stavelsen, og tonemarkeringen fungerer samtidig som markering av hovedtrykk. Også hovedtrykk der tonem er nøytralisert, markeres med superskriberte tall, nærmere bestemt 1-tall. Bitrykk, stavelsesgrense eller morfemgrense markeres normalt ikke, med mindre disse fenomenene blir diskutert eksplisitt i den aktuelle sammenhengen. Bitrykk markeres med subskriberte 1-tall.

1.2.3 Forkortelser

Følgende ordbøker er referert til med forkortelser:

- | | | |
|-------------------------------|------|------------------------------|
| • <i>Norsk Uttaleordbok</i> | NUOB | (Berulfsen 1969) |
| • <i>Norsk Riksmålsordbok</i> | NROB | (Knudsen og Sommerfelt 1983) |
| • <i>Bokmålsordboka</i> | BMOB | (Landrø og Wangensteen 1993) |
| • <i>Norsk Uttaleordbok</i> | NUOV | (Vanvik 1985) |

Følgende nettordbøker er referert til:

- | | |
|-----------------------------------|---|
| • <i>Bokmålsordboka</i> | http://www.dokpro.uio.no/ordboksoek.html |
| • <i>Norsk Riksmålsordbok</i> | http://www.ordnett.no/ordbok.html |
| • <i>Nynorsk frekvensordliste</i> | http://www.tekstlab.uio.no/ |

De brukte korpus er referert til på følgende måte:

- Oslo-korpuset <http://www.tekstlab.uio.no/norsk/bokmaal/>
- Usenet-korpuset <http://logos.uio.no/usenet.html>

Andre forkortelser:

- SØ Standard østnorsk

1.3 Metode og datagrunnlag

1.3.1 Ordbøker

Forskningsobjektet for denne oppgaven er selve bokmålsnormen, eller en del av den, se kapittel 2.2.2. Det er altså ikke *bruken* av denne normen, *usus*, som er forskningsobjektet.¹ Som primært datagrunnlag har jeg brukt en papirutgave av *Bokmålsordboka* (BMOB²), supplert med nettutgaven³ og en tekstfil som utgjør deler av grunnlaget for produksjonen av boka, og som inneholder alle bøyningsformer og alle stavemåter av alle ord som står i *Bokmålsordboka*. Disse tre kildene er ikke fullstendig overensstemmende, delvis på grunn av det siste tiårs justeringer av normen, men dette er i liten grad et praktisk problem for oppgaven.

Nettutgaven har jeg aksessert med den søkefunksjonaliteten som er tilgjengelig på nettsiden, og som har sterke begrensninger, både hva gjelder søkefunksjonalitet og presentasjonsmuligheter. For systematiske søk i ordbasen har jeg derfor benyttet meg av søk med *regulære uttrykk* i en tekstfil som inneholder alle ordene i ordboka. Filen består av én linje for hver bøyningsform, og hver linje består av kategoriinformasjon i tillegg til leksemet på leksikonform og en bøyningsform. I tillegg finnes enkelte tall- og bokstavkoder som benyttes i produksjonen av filens egentlige formål. Her er et utdrag:

¹Se likevel 1.3.2 på side 15.

²Landrø og Wangensteen 1993.

³<http://www.dokpro.uio.no/ordboksoek.html>

subst	mask	appell	bretonsk	bretonskene	8827	4
subst	mask	appell	brett	brett	8828	1
subst	mask	appell	brett	bretten	8828	2
subst	mask	appell	brett	bretter	8828	3
subst	mask	appell	brett	brettene	8828	4
subst	mask	appell	brettsegler	brettsegler	8837	1
subst	mask	appell	brettsegler	brettsegler	8837	1
subst	mask	appell	brettsegler	brettsegleren	8837	2
subst	mask	appell	brettsegler	brettsegleren	8837	2
subst	mask	appell	brettsegler	brettseglere	8837	3
subst	mask	appell	brettsegler	brettseglerer	8837	3
subst	mask	appell	brettsegler	brettseglerne	8837	4
subst	mask	appell	brettsegler	brettseglerne	8837	4
subst	mask	appell	brettsegling	brettsegling	8838	1
subst	mask	appell	brettsegling	brettseglingen	8838	2
subst	mask	appell	brettsegling	brettseglinger	8838	3
subst	mask	appell	brettsegling	brettseglingene	8838	4
subst	mask	appell	brettseiler	brettseiler	8840	1
subst	mask	appell	brettseiler	brettseiler	8840	1
subst	mask	appell	brettseiler	brettseileren	8840	2
subst	mask	appell	brettseiler	brettseileren	8840	2
subst	mask	appell	brettseiler	brettseilere	8840	3
subst	mask	appell	brettseiler	brettseilerer	8840	3
subst	mask	appell	brettseiler	brettseilerne	8840	4
subst	mask	appell	brettseiler	brettseilerne	8840	4
subst	mask	appell	brettseiling	brettseiling	8841	1
subst	mask	appell	brettseiling	brettseilingen	8841	2
subst	mask	appell	brettseiling	brettseilinger	8841	3
subst	mask	appell	brettseiling	brettseilingene	8841	4
subst	mask	appell	brevadel	brevadel	8845	1

Som det går fram av utdraget over, får hver eneste tillatte bøyningsform minst én linje, og alle mulige varianter av grunnformen må kombineres med alle mulige varianter av alle relevante bøyningsmorfemer. Vi ser at alle ulike bøyningsformer og valgfrie former gjør tekstfilen temmelig voluminøs. Faktisk er hele filen på nesten 30 megabyte, og dette er altså kun tekst! Teksten utgjør godt over 700.000 linjer, noe som skulle tilsi minst 10.000 sider om det skulle trykkes på papir.

For å søke i denne tekstmengden har jeg benyttet et program av grep-familien, som kan søke i tekst ved hjelp av regulære uttrykk. I denne sammenhengen er det en fordel at hver bøyningsform er presentert på hver sin linje, men det er til tider en ulempe at alle bøyningsformer er med i det hele tatt. Heldigvis er linjen som inneholder grunnformen, markert med et 1-tall helt på slutten av linjen, og denne informasjonen har jeg brukt i de regulære uttrykkene for å avgrense antall treff.

Vi ser videre at det ikke finnes junkturmarkeringer i listen, noe som ytterligere vanskeliggjør søkingen og resulterer i mye høyere trefffrekvens enn ønskelig. Problemet er todelt, men kan illustreres med ett eksempel:

Jeg ønsket å undersøke hvilke ord som inneholdt sekvensen <kkj>, noe som kunne gjøres ved følgende regulære uttrykk:

```
\s.*[aeiouyæøå]kkj[aeiouyæøå].*\s.*[^\d]1\r\n
```

Søket treffer sekvenser av <kkj> med minst én vokal på hver side.¹ Den delen av søkeuttrykket som kommer etter den siste asterisken, krever at det skal komme et enslig 1-tall helt til slutt på den linjen som inneholder grafemsekvensen, altså at vi bare er interessert i linjene som bare inneholder grunnformen. Resultatet av dette søket ble 23 treff:

1	fem	appell	bikkje	bikkje	6859	1
1	fem	appell	brakkjord	brakkjord	8514	1
1	fem	appell	fillebikkje	fillebikkje	16930	1
1	fem	appell	gneldrebikkje	gneldrebikkje	23805	1
1	fem	appell	krykkje	krykkje	38045	1
1	fem	appell	lausbikkje	lausbikkje	39909	1
1	fem	appell	løsbikkje	løsbikkje	42663	1
1	mask	appell	bikkje	bikkje	6859	1
1	mask	appell	blekkjertel	blekkjertel	7488	1
1	mask	appell	brakkjord	brakkjord	8514	1
1	mask	appell	butikkjede	butikkjede	9827	1
2	mask	appell	fabrikkjører	fabrikkjører	15728	1
1	mask	appell	fillebikkje	fillebikkje	16930	1
1	mask	appell	gneldrebikkje	gneldrebikkje	23805	1
1	mask	appell	krykkje	krykkje	38045	1
1	mask	appell	lausbikkje	lausbikkje	39909	1
1	mask	appell	løsbikkje	løsbikkje	42663	1
2	mask	appell	møkkjeller	møkkjeller	46787	1
2	nøyt	appell	bikkjeleven	bikkjeleven	6861	1
2	nøyt	appell	bikkjeslagsmål	bikkjeslagsmål	6862	1
2	nøyt	appell	brekkjern	brekkjern	8745	1
1			bikkjekald	bikkjekald	6860	1
1			skakkjørt	skakkjørt	60451	1

Av disse treffene var jeg egentlig interessert i kun 2: <bikkje> og <krykkje>. De resterende havner i tre kategorier:

¹I dette tilfellet reduserte kravet om vokaler på hver side av den aktuelle grafemsekvensen antall treff kun med to: <atlaskkjole> og <skjoldbruskkjertel>, men i andre tilfeller vil en slik presisering av søket være til større hjelp for å kutte ned på antall irrelevante treff.

1. sammensetninger av formen <kk>+<j> eller <k>+<kj>¹.
2. sammensetninger med *bikkje* som enten første- eller andreledd, f.eks. *bikkjekald* eller *fillebikkje*.²
3. *bikkje* er listet som oppslagsord to ganger, én for hvert mulig genus – *bikkja* eller *bikkjen*.

Punkt 1 inneholder selvfølgelig potensielt nyttig informasjon, men jeg var klar over at det fantes slike sammensetninger, og jeg kjente til nok eksempler av denne typen til at jeg kunne illustrere et slikt poeng. Målet med søket var å finne intervokaliske forekomster av den aktuelle konsonantsekvensen innenfor ett morfem, og med mulighet for å spesifisere at søket ikke skulle treffe forekomster av <kkj> som inneholdt en morfemgrense, kunne jeg ha redusert antall treff med 8, eller 35%.

Resultatene i punkt 2 inneholder sjelden interessant informasjon. Om jeg finner en stamme som tilfredsstiller søkekriteriene mine, er det ikke overraskende at det også finnes sammensetninger der denne stammen inngår, og også disse treffene gjør resultatlisten mindre oversiktlig. Ved å fjerne alle sammensetningene med <bikkje> fra resultatet, ville resultatene ha inneholdt 11 færre treff, altså 48%. Enda bedre ville det ha vært om jeg kunne ha spesifisert søket til å treffe kun røtter i utgangspunktet. Da ville jeg ha stått igjen med kun 3 treff, bare 1 mer enn jeg var interessert i.

Punkt 3 kunne jeg ha unngått ved å spesifisere at jeg bare ønsket treff på ett genus for hvert leksem, men dette ville ha vært et vesentlige tyngre søk med regulære uttrykk. I dette tilfellet økte det bare antall treff med 7, men det utgjør faktisk 30% av treffene og representerer et større praktisk problem med et antall treff på f.eks. 230 eller 2300.

¹Faktisk er det ingen tilfeller av typen <k>+<kj> men bare av typen <kk>+<kj> → <kkj>.

²Sammensetningene illustrerer for øvrig godt påstanden i *Bokmålsordboka* om at ordet *bikkje* har negative konnotasjoner!

Som vi ser, medfører søking med regulære uttrykk i store tekstfiler ofte et behov for å etterbehandle resultatlistene manuelt. I noen tilfeller blir antall treff så høyt at det blir praktisk umulig å etterbehandle treffene manuelt, og det blir dermed i realiteten umulig å utføre søket. Søking med regulære uttrykk i tekstfilen gir likevel store muligheter til systematiske undersøkelser av bokmålsnormen som ville ha vært praktisk umulige ved hjelp av den trykte boka eller nettstedet slik det framstår i dag.

I tillegg til *Bokmålsordboka* har jeg dessuten i noen grad benyttet *Norsk Riksmålsordbok* (NROB¹), til informasjon om både uttale, staving, grammatiske opplysninger og etymologi. *Norsk Riksmålsordbok* bygger ikke på samme norm som *Bokmålsordboka*, men med kjennskap til tendensene i de ulike syn på rettskrivning som er representert i de to bøkene, er det som oftest mulig å benytte *Norsk Riksmålsordbok* som rettesnor for ord som ikke finnes i *Bokmålsordboka*. *Norsk Riksmålsordbok* har et vesentlig større omfang enn *Bokmålsordboka* og har derfor en viss nytteverdi for oppslag om lavfrekvente ord. Jeg har også benyttet nettutgaven av *Norsk Riksmålsordbok*².

1.3.2 Korpus

I tillegg til ordbøker har jeg benyttet to elektroniske korpus til datainnsamling. Som nevnt over, er ikke formålet med denne oppgaven primært å undersøke hvordan språkbrukere bruker bokmålsnormen, men studier av korpus har likevel en nytteverdi, på to ulike måter.

For det første ønsket jeg å undersøke tendenser til statistisk betingede ulikheter i normen. Ved hjelp av korpus kunne jeg også undersøke hvorvidt enkeltstående tilfeller eller enkeltprinsipper i den kodifiserte normen er marginale med hensyn til usus og dermed den operative normen.

¹Knudsen og Sommerfelt 1983.

²<http://www.ordnett.no/ordbok.html>

For det andre kunne jeg bruke korpus til å teste hypoteser om sammenheng mellom valgfrie stavemåter og talemålsnormer.

Dessuten er kunnskap om usus av interesse i visse tilfeller, for eksempel i et normeringsperspektiv, selv om studier av usus ikke er noe sentralt punkt i oppgaven. Eksempler på interessante problemstillinger er:

- Brukes begge stavemåter der det er valgmuligheter?
- Brukes bokmålsnormen eller brukes heller riksmålsnormen i visse tilfeller?
- Kjenner språkbrukerne normen i alle tilfeller – f.eks. prinsipper som i hvert fall tilsynelatende bryter med hovedprinsippene i normen?

Jeg har konsultert to korpus, det såkalte *Oslo-korpuset* og et korpus av tekster fra diskusjonsgruppene på Usenet, som jeg kaller *Usenet-korpuset*. Nettadresser for disse er gitt i 1.2.3.

Oslo-korpuset er satt sammen av Tekstlaboratoriet ved Universitetet i Oslo. Det består av bokmålstekster på til sammen omtrent 18,3 millioner ord. Tekstene er fordelt mellom drøyt 50% avis- og ukebladtekster, snaut 40% sakprosaetekster og ca. 10% skjønnlitterære tekster. Sakprosaetekstene er utelukkende offentlige utredninger, lover og forskrifter. Alle tekstene er fra før år 2000. Det er verdt å merke seg at selv om 18,3 millioner ord umiddelbart virker som en stor tekst-samling, har den visse begrensninger. Utvalget er ikke større enn at en del lavfrekvente ord i liten eller ingen grad er representert. Om et søk var ment å skille mellom høyfrekvente og lavfrekvente ord, er ikke dette noe problem i seg selv, men om søket skulle belyse språkbrukernes intuisjon eller usus, er det et problem. Tekstutvalget er heller ikke på noen måte representativt for skriftlig språkbruk på norsk; blant annet er utvalget av "sakprosaetekster" svært ensidig, og utvalget mangler også mer personlig pregede tekster. Utvalget inneholder også en del tekster på nynorsk, noe som til en viss grad vanskeliggjør tolkningen av enkelte søk, særlig i de tilfellene et lavfrekvent bokmålsord er homograft med et

høyfrekvent nynorskord,¹ for eksempel nynorske flertallsformer for hannkjønnsord som *sonar, pulsar, radar*.

Når det gjelder søkemuligheter i korpuset, er de gode. Det finnes et søkegrensesnitt med søkefelt, rullegardinsvalg, avkrysningsbokser og knapper som gir et svært kraftig grep om samlingen, og det er dessuten mulig å søke med regulære uttrykk til de få formål der det spesialutviklede søkegrensesnittet må melde pass.

Tekstene er dessuten "tagget" med grammatisk informasjon som ordklasse og bøyningsform. Taggingen er utført automatisk av et eget taggeprogram, som ifølge Tekstlaboratoriet har en suksessfrekvens på 99%. Dessverre er ikke programmet kraftig nok til å skille konsekvent mellom homografer av ulike ordklasser, slik som <leser>,v=/¹le:ser/ : <leser>,s=/²le:ser/, men tilegner en slik ordform tagger for begge ordklasser. Vi kan derfor ikke bruke Oslo-korpuset til å undersøke relative bruksfrekvenser for homografpar.

Noen av disse problemene utlignes av Usenet-korpuset, som i prinsippet består av alle tekster under no.*-hierarkiet på Usenet i perioden 1998-2002, totalt ca. 140 millioner ord¹. Det er med andre ord vesentlig større – en størrelsesorden over – og det supplerer i tillegg sjangerutvalget fra Oslo-korpuset ved at det inneholder sakprosa tekster fra andre fag og sjangere. Det inneholder dessuten tekster av mer personlig karakter, og det er viktig å merke seg at til forskjell fra for tekstene i Oslo-korpuset er de fleste av skribentene som er representert i Usenet-tekstene, ikke profesjonelle skribenter.

En annen viktig forskjell i forhold til Oslo-korpusets tekster er at Usenet er mindre formelt, på flere måter:

¹Avisdelen av korpuset inneholder 976 forekomster av <ikkje>, som er et ganske sikkert kriterium for at teksten er nynorsk. De fleste av forekomstene stammer fra Bergens Tidende og Vårt Land. (Det samme kunne ikke med like stor sikkerhet hevdes om skjønnlitterære tekster, der forekomster av <ikkje> kunne stamme fra blant annet dialoger gjengitt på dialekt i en ellers ren bokmålstekst.

¹Nettsiden opplyser at enkelte innlegg av tekniske årsaker ikke er kommet med. Sitater fra andre innlegg er fjernet i den grad det er teknisk mulig å gjøre det automatisk.

- Det finnes ingen korrektur som retter opp grammatiske feil, stavefeil eller trykkfeil.
- Det finnes ingen redaktør som redigerer eller forkaster tekster av dårlig kvalitet eller liten relevans.¹
- Tekstene er gjerne preget av en mer muntlig stil, både i ordvalg og i setningsbygning. Noen innlegg er også svært korte. Det er sannsynlig at den muntlige stilen også kan påvirke stavingen til skribentene.

Som en følge av de to første punktene over, og som et uttrykk for det siste, er det liten tvil om at mange innlegg på Usenet blir produsert i svært høy hastighet og med liten grad av korrektur eller språkkontroll fra skribentenes side.

Søkegrensesnittet mot Usenet-korpuset er en forenklet utgave av grensesnittet mot Oslo-korpuset, og det mangler fullstendig muligheten for søking med regulære uttrykk. I praksis gir dette likevel få begrensninger.

1.3.3 Frekvensordlister

Et viktig biprodukt av korpusene er frekvensordlister. Av Oslo-korpuset er det produsert en frekvensordliste av de 10000 mest frekvente ordene i avis- og ukebladtekstene. Ordlisten er presentert på en egen nettside² der ordlisten er delt oppi bolker på 1000 ord hver. Ordlisten er dermed ikke søkbar i seg selv, men ved å laste ned alle delene og importere dem i et regneark, får man en del enkle søkings- og sorteringsmekanismer. Om man trenger mer avanserte metoder, kan ekstra funksjonalitet programmeres, f.eks. i OpenOffice.org Basic, som er integrert med regnearkprogrammet OpenOffice Calc. Jeg genererte en tekstfil og brukte `grep`-programmet som er omtalt over, til å søke i ordlisten.

¹Noen grupper er såkalt "modererte"; det vil si at det faktisk finnes en redaktør som godkjenner eller forkaster innlegg på grunnlag av innholdets kvalitet og/eller relevans. Selv om slike redaktører neppe redigerer det språklige i innleggene, er det lett å tenke seg at innsenderne til slike grupper i større grad legger vekt på det språklige og bruker mer tid på gjennomlesning og korrektur før publisering.

²http://www.tekstlab.uio.no/norsk/frekvensordlister/bokmaal/aviser_og_lover/aviser.frek.html

Jeg brukte frekvensordlista som et alternativ til *Bokmålsordboka*. Ved å søke opp relevante ordformer fra frekvensordlista har jeg unngått en del av de problemene ved å søke i ordboksfila som jeg omtalte over, og fått en større andel relevante treff. For eksempel fikk jeg ved søk i ordboksfilen etter ord som inneholdt grafemsekvensen <gn>, 10466 treff, mens jeg i frekvensordlista fikk 78, som utgjorde et passe utvalg av ordformer som jeg faktisk vet er i bruk.

Det er viktig å merke seg at frekvensordlisten fra Oslo-korpuset er generert ut fra et enda snevrere utvalg av tekstsjangere enn hele korpuset.

En annen frekvensordliste som er gjort tilgjengelig via nettsidene til Tekstlaboratoriet, er *Nynorsk frekvensordliste*, som opprinnelig er en papirproduksjon (Vestbøstad 1989). Nettutgaven består av de 19369 mest frekvente ordene fra tekster fra dags- og lokalaviser (50%), ukeaviser (30%) og skjønnlitteratur (20%), og presentasjonen er den samme som for frekvensordlisten fra Oslo-korpuset og gir dermed de samme mulighetene for nedlasting for videre manipulering i andre programmer.

1.3.4 Uttaleordbøker

Jeg er kjent med at det finnes i hvert fall utkast eller beta-utgaver av elektroniske uttaleordbøker for norsk for kommersielle formål. Det har ikke lyktes meg å få tilgang på noen slik ressurs. Det er klart at enkelte av undersøkelsene mine kunne ha vært gjort vesentlig enklere med tilgang på et slikt materiale – enklere i en så stor grad at det kunne ha blitt en helt annen undersøkelse.

I mangel av andre kilder har jeg benyttet meg av de to nyeste trykte uttaleordbøkene for norsk, nemlig *Norsk uttaleordbok* fra 1969, redigert av Bjarne Berulfsen (NUOB), og en bok med samme navn fra 1985, redigert av Arne Vanvik (NUOV). I tillegg finnes sporadisk uttaleinformasjon i *Bokmålsordboka* og i *Norsk Riksmålsordbok*, som jeg har benyttet meg av, ved siden av egen intuisjon.¹

¹Jeg er født i 1963 og vokst opp på Hamar, men med foreldre fra Oslo. Jeg er nå bosatt i nærheten av Hamar.

Berulfsens uttaleordbok har visse svakheter. For det første begynner den å bli noe gammel; når denne hovedoppgaven blir skrevet, er ordboken 36 år gammel. Dertil kommer at redaktøren på det tidspunkt ordboken ble utgitt, allerede var 63 år. Det er med andre ord ikke usannsynlig at den talemålsnorm som blir presentert, ligger et stykke fra dagens.

For det andre analyserer den de retroflekse konsonantene som sekvenser av ikke-retroflekser, slik at ord som <fort>, <ferdig>, <barn>, <Karlsøy>, <særs> blir transskribert som /fort/, /²færdi/, /ba:rn/, /¹ka:rlsøy/, /sæ:rs/.¹ Dette er ikke i tråd med analysene i f.eks. Haslev 1985, Sivertsen 1967, Vanvik 1979 eller Kristoffersen 2000, som alle analyserer retrofleksene som egne fonemer.² Berulfsens analyse fører til problemer. Det er nemlig ikke slik at det er fullstendig automatikk i assimilasjonen av /r/ og dentalene i alle omgivelser. Særlig /r/+d/ har mye variasjon i utbredelsen av denne assimilasjonen, og i standard østnorsk er den blokkert i ord som <verden>=²værdn/ og <sverd>=²sværd/.³ Dette har dessuten innvirkning på kvantiteten i visse ord; *ferdig* uttalt med retrofleks kan nemlig også ha lang vokal, noe som ikke går fram av NUOB.

Berulfsen skiller mellom [e] og [ə] i for eksempel <felle>=[2felə] og mellom [ø] og [ö] i for eksempel <øs>=[ø:s] og <øst>=[öst], men han skiller ikke mellom [e] og [ɛ] i for eksempel <fet>=[fe:t] og <fett>=[fɛ]. Det er liten tvil om at det er markert *fonetisk* forskjell mellom alle disse tre parene, men om det er *fonemisk* distinksjon, er gjenstand for diskusjon. Berulfsen indikerer ikke om transskripsjonene representerer fonemiske eller fonetiske former, noe som ville ha vært av betydning for vår bruk av disse dataene. Vanvik kommenterer i sitt forord

¹Berulfsen bruker andre transskripsjonskonvensjoner. Jeg har her transskribert Berulfsens lydskrift til den som ellers er i bruk i denne oppgaven.

²Borgstrøm 1958 gjør det imidlertid ikke.

³Mine egne observasjoner av barn i hvert fall på Hedmarken tyder på at det kan være bevegelse i normen her. De sier /¹vædn/ og /svædn/. Tiden vil vise om denne uttalen vil følge dem inn i voksenormen.

at Berulfsens uttaleordbok nok må være den eneste i verden som er redigert av en ikke-fonetiker, og jeg ser ikke bort fra at Vanvik da han skrev dette, kan ha hatt nettopp problemet med den uklare fonemiske statusen i tankene.

Vanviks uttaleordbok lider ikke av noen av disse problemene. Den er nyere, Vanvik er fonetiker, og selv om han ikke sier det uttrykkelig, er det klart at han etterstreber en fonemisk transskripsjon. Han bruker retroflekse konsonanter, og han lister også opp sammentrukne former som

<størrelse>=/²stø:lse/.¹

Formene i Vanviks uttaleordbok stemmer i svært stor grad overens med min egen intuisjon. Der det er avvik, har jeg påpekt det eksplisitt, dersom jeg har hatt sterk nok tillit til min egen intuisjon.

Begge disse uttaleordbøkene har imidlertid en mangel som vedkommer mitt formål ved bruken av dem; de lister ikke opp uttalen av bøyde former. Dette er et metodisk problem og fører til en del tilfeller der jeg utelukkende må stole på egen intuisjon.²

1.4 Resten av oppgaven

Kapittel 2 diskuterer avgrensningen av oppgavens problemstilling.

Kapittel 3 gir en kort oversikt over de relevante delene av språkssystemet i standard østnorsk. Noen aspekter ved språkssystemet er diskutert mer inngående i sammenheng med omtale av de relaterte delene av skriftsystemet lenger ut i oppgaven.

Kapittel 4 refererer, introduserer og diskuterer grunnleggende skrifteori som er nødvendig for å analysere rettskrivningen i bokmål. Teorien er hovedsakelig

¹Berulfsen har her bare /²størelse/.

²Et eksempel er hvorvidt perfektumformen av *leke* kan ha både lang og kort vokal. Dette eksemplet er brukt som illustrasjon i kapittel 3.2.

illustrert av eksempler fra bokmål, slik at en vesentlig del av analysen av bokmål finner sted i dette kapitlet, parallelt med omtalen av teorien.

Kapittel 5 gir en oversikt over variasjonen i bokmål sett fra et fonemografisk perspektiv, og det danner dermed et datagrunnlag for mye av diskusjonen i kapittel 4. Dessuten blir momenter som ikke er berørt i det teorikapitlet, behandlet her.

Kapittel 6 viser andre synsvinkler å studere bokmål rettskrivning fra.

Kapittel 7 omtaler kvantitative undersøkelser av bokmål gjennom analyser av utdrag av standard østnorsk leksikon og studier av Oslo-korpuset.

Til slutt i oppgaven er det en referanseliste og stikkordsregister, samt en kort liste over etymologier for importord som er omtalt i oppgaven.

2 Avgrensninger

Denne oppgaven omhandler bokmålets internsystemiske egenskaper og forholdet mellom skrift og tale på fonem-, morfem- og ordnivå. Formulert på denne måten er dette et svært stort område som omfatter mange ulike problemstillinger. Dette kapitlet diskuterer avgrensningene av problemområdet.

2.1 Standard østnorsk

2.1.1 Begrepet

Oppgaven omhandler bokmålets forhold til taleformer som tilhører standard østnorsk (SØ). Begrepet *standard østnorsk* er imidlertid ikke noen term med avklart innhold.

Først og fremst må det sies at heller ikke selve betegnelsen *standard østnorsk* er allment akseptert. Berulfsen 1969:3-4 diskuterer begrepet *dannet talemål* og påpeker at dette er blitt mer tøyelig i Norge enn i andre land, og at det i Norge i stedet for én riksnorm finnes flere "så å si lokalrikspråk", men han etterstreber i uttaleordboken en transskripsjon som gjenspeiler "den normaluttale som stort sett er rådende blant folk med høyere allmenndannelse i hovedstaden og andre byer på Østlandet".

Kristoffersen 2000:8-9 skriver at de fleste tidligere verk som omhandler norsk fonologi, enten eksplisitt eller implisitt har omhandlet SØ. Kristoffersen påpeker at SØ ikke er offisielt kodifisert, og de vanskeligheter som oppstår som en konsekvens av dette. Han hevder at det er vanskelig å skille hovedstadens standardtalemål fra middelklassens talemål i andre østnorske byer, og han bruker derfor begrepet *urban østnorsk* om de talemålsformene han omtaler. Han viser dessuten til den stilvariasjonen som talere av urban østnorsk har tilgang til, særlig med hensyn til variasjon mellom ulike bøyingsmorfemer og mellom retrofleks flapp ("tjukk l") og lateral – [ɾ] og [l]. Variasjonen mellom bøyingsmorfemer

angår ikke min problemstilling, men den nevnte fonologiske variasjonen gjør det til en viss grad.

Vanviks *Norsk uttaleordbok*, som er oppgavens viktigste datagrunnlag på uttalesiden, er blant de behandlinger av norsk talemål som er eksplisitt i sin omtale av SØ. Han definerer SØ som "prestisjeuttalen i den sørøstlige delen av landet med hovedstaden som det naturlige sentrum. Sosioøkonomisk sett brukes SØ gjerne av velutdannede personer [...]", men han påpeker også at SØ først og fremst er en geografisk og i mindre grad en sosioøkonomisk varietet, i motsetning til standardtalemål i mange andre land, for eksempel England. I forhold til Oslo-området mer "folkelige" uttale, utgjøres de største forskjellene av distribusjonen av retrofleks flapp og trykkplassering i en del lånord.

I tillegg til en geografisk og en sosial avgrensning av SØ, er det nødvendig med en temporal avgrensning. Vanvik påpeker for eksempel nyere utviklinger i SØ, blant annet senkningen av startpunktet i diftongen som vanligvis gjengis som <ei>, altså i retning fra [e] mot [æ]. For vårt formål er det tilstrekkelig å si at vi studerer bokmålets forhold til *dagens* SØ. Detaljene i språkssystemet i SØ blir behandlet i mer detalj i kapittel 3.

2.1.2 Historien

Den historiske bakgrunnen for SØ skal ikke behandles her, men det skal bare påpekes at dagens SØ i stor grad er blitt til i et samspill mellom norskfødte danskers norskpregede uttale av dansk språk i Oslo på den ene siden, og den opprinnelige vikkværske dialekten på den andre. Mer detaljerte utlegninger om dette finnes blant annet i Kristoffersen 2000 og Torp og Vikør 2003.

2.2 Bokmål

2.2.1 Historisk bakgrunn

Store deler av de siste tusen år av norsk språkhistorie er relativt godt kjent (se for eksempel Torp og Vikør 2003), både når det gjelder skriftspråsutviklingen og talespråkutviklingen, som også er relevant i en diskusjon av bokmålet.

Talespråkutviklingen har gått fra norrønt via mellomnorsk til moderne norsk, i hovedstadsområdet med sterk påvirkning fra dansk gjennom århundrer med dansk styre. Skriftspråkutviklingen har etter å ha beveget seg fra runeskrift til latinske bokstaver vært aller mest preget av bruk av dansk skriftspråk etterfulgt av en fornorskningssprosess de siste 150 år eller så.

Med en grov forenkling kan vi si at foruten gjennom den indre talemålsutviklingen har dagens bokmål vært utsatt for seks viktige eksterne påvirkningsfaktorer: i kronologisk rekkefølge norrønt, nederlandsk og (neder)tyisk, dansk, latin og gresk, nynorsk og engelsk.

Å kalle norrønt språk for en *ekstern* påvirkningsfaktor kan selvfølgelig argumenteres mot, men i rettskrivningssammenheng kan den kalles ekstern i den forstand at norrøne rettskrivningskonvensjoner opererte på et språkstadium som er så ulikt SØ at det ifølge et "gjensidig forståelse"-kriterium må kalles et annet språk. Likevel kan spor av flere av disse konvensjonene finnes igjen i dagens bokmål. Fordelingen av <u o å> over fonemene /u o å/ er et åpenbart eksempel, der <o>, <u> representerer fonemene /å/, /o/ i stedet for primærfonemene /o/, /u/ etter et mønster som i stor grad følger norrøn rettskrivning.¹

SØ og bokmål inneholder en svært stor andel av importord av nederlandsk, tysk eller nedertysk opprinnelse (Sandøy 2000). Mange av disse ordene skiller seg ikke i dag fonologisk eller fonotaktisk fra ord av eldre norsk opprinnelse, men en del av dem gjør det, og disse har bidratt til det systemet av regler og tendenser som styrer bokmål rettskrivning i dag. Velkjente eksempler er ord med frontale,

¹Dette er dokumentert i min egen upubliserte grunnfagsoppgave fra 1996 og omtalt i Torp og Vikør 2003.

trykklette avledningsmorfemer som *be-* og *ge-*, som har medvirket til å endre de fonotaktiske reglene for trykkplassering i SØ-ord, og frontale, trykksterke avledningsmorfemer som *an-* og *fore-*, som avviker fra hver sin dominerende regel for bokmål:

- kort vokal i trykksterk stavelse lisensierer etterfølgende dobbeltkonsonant i samme morfem, som f.eks. i *inn-*.
- /å:/ lisensierer <å> med visse unntak, men unntakene er særlig foran <v> og <g> (ifølge Eitrem og Sandvei 1960).

Foruten å ha satt sitt preg på talemålet har dansk også satt spor etter seg i norske skriftspråskkonvensjoner, på tross av en langvarig fornorskningsprosess etter dansketida.

Gjennom hele vår kjente språkhistorie har språket vårt blitt påvirket av de "klassiske" språk, altså gresk og latin, både gjennom direkte import og gjennom nyskapninger dannet på klassiske stammer. Import av ord eller stammer fra disse språkene har i stor grad beholdt i hvert fall deler av rettskrivningen fra kildespråket, også der disse strider mot bokmålets prinsipper og mønstre, og selv om ordene er veltilpasset både fonologisk og morfologisk. Et eksempel er det latinske *transcribo*, som er en sammensetning av *trans* + *scribo*, men der en gammel latinsk rettskrivningsregel sørger for reduksjon av den doble <s> i fugen. Dette mønsteret er beholdt i bokmål, til tross for at det er fremmed for andre grupper av ord i bokmål og ville ha produsert en form som *<avik> av *av* + *vik*.¹

I vesentlige deler av forrige århundre var det en uttalt politikk å bidra til at Norges to offisielle målformer landsmål/nynorsk og riksmål/bokmål nærmet seg hverandre. Dette innebar både at endringer ble foretatt i bokmålsrettskrivningen for å bringe den nærmere nynorsk, og at endringer i nynorsknormen medførte parallelle endringer i bokmålsrettskrivningen, selv der disse var mindre systemrette i bokmål. Mange av disse endringene dreide seg om endringer i

¹<transkribere>, <transskribere> og <transskripsjon>, <transskripsjon> er riktig nok sidestilte former, men Usenet-korpuset viser at formene med én <s> er dominerende, med ca. 90% av forekomstene. Jeg velger å følge bokmålmønsteret og skriver disse ordene med to <s>-er i denne oppgaven.

hvilke bøyningsmorfemer i talespråket som skulle kunne gjengis i rettskrivningen, og slike problemstillinger er ikke tema for denne oppgaven, men noen dreide seg også om endringer som vedkommer forholdet mellom skriftbilde og et udiskutabelt talemålsgrunnlag, for eksempel fjerning av <h> fra <hveps> og <erhverv> i 1938 og <hverken> og <hvalp> i 1959. Slik kan vi si at nynorsk har vært en ekstern påvirkningsfaktor.

Særlig de siste tiårene har påvirkningen fra engelsk vært merkbar og omdiskutert.¹ Det engelske skriftsystemet er temmelig ulikt bokmål, og det er i stor grad dette som gjør at import av engelske ord er konfliktfylt, men også forskjeller i fonologi og fonotaks spiller en rolle her.

2.2.2 Kodifisert og faktisk norm²

Oppgaven omhandler forholdet mellom SØ og bokmål.

Bokmålsnormen er, slik den har utviklet seg gjennom departementsvedtak det siste hundreåret, en vid norm med mange valgmuligheter og endog valgmuligheter av to ulike grader.³ Oppgavens problemstilling er imidlertid ikke å vurdere forholdet mellom de ulike valgmulighetene eller i hvilken grad de gjenspeiler talemålsformer. Oppgavens problemstilling er å finne de skriftmålsformer som representerer talemålsformer som er en del av SØ.

1. <segl> og <seil> er begge med i bokmålsnormen, og de representerer begge talemålsformen /sæjl/⁴. Både <segl>=/sæjl/ og <seil>=/sæjl/ er korrespondanser som er aktuelle å diskutere.
2. <trappa> og <trappen> er begge med i bokmålsnormen, og de representerer henholdsvis /^ltrapa/ og /^ltrapen/, som begge kan hevdes å være med i SØ,

¹Se for eksempel Sandøy 2000.

²Dyvik 2003 bruker termene *foreskrevne normer* og *operative normer*. Artikkelen står i Omdal og Røstad 2003.

³I skrivende stund ser det ut til at systemet med to ulike grader av valgmulighet blir avskaffet.

⁴Om den fonemiske transskripsjonen som blir benyttet i oppgaven, se 1.2.2.

dog med ulikt stilnivå. Både <trapa>=¹trapa/ og <trappen>=¹trapen/ er korrespondanser som er aktuelle å diskutere, mens en vurdering av de to grafiske formene opp mot hverandre faller utenfor oppgaven, og en diskusjon av en eventuell korrespondanse <trappa>*=¹trapen/ er uaktuell.

3. <kull> og <kol> er begge med i bokmålsnormen. <kull> representerer talemålsformen /kul/, mens <kol> kanskje kan sies å korrespondere med /kå:l/, /kå:ɾ/ eller /kø:ɾ/. Ingen av disse fonologiske formene kan imidlertid sies å falle inn under SØ; de to første faller utenfor de geografiske avgrensningene, mens den siste faller utenfor de sosiale avgrensningene og er del av det Vanvik kaller "folkelig" talemål. <kull>=/kul/ er dermed med i diskusjonen, mens <kol> ikke er det.
4. <snøggto> og <hurtigto> er begge med i bokmålsnormen og representerer de fonologiske formene ²/snøgtå:g/ og ²/huɽitå:g/. Selv om det er liten tvil om hvilken uttale som ville bli produsert av en SØ-språkbruker som ble tvunget til å lese ordet <snøggto>, er dette ordet ikke i bruk i SØ. <snøggto>=²/snøgtå:g/ er dermed ikke med i diskusjonen, mens <hurtigto>=²/huɽitå:g/ er.
5. /sne:/ og /snø:/ er begge del av SØ, og de kan representeres i skrift med henholdsvis <sne> og <snø>, og igjen er det ingen tvil om hvilken uttale som vil bli realisert ved høytlesing av disse ordformene av en SØ-språkbruker. Imidlertid er ikke <sne> del av offisiell bokmålsnorm, og <sne>=/sne:/ faller derfor utenfor diskusjonen, mens <snø>=/snø:/ faller innenfor.¹

Vi ser av eksemplene over at vi ønsker å utelukke skriftformer av flere ulike årsaker:

- skriftord som ikke er del av bokmålsnormen
- skriftord som ikke representerer en SØ-uttale
- skriftord som ikke representerer et SØ-ord

¹Jeg vil likevel i enkelttilfeller diskutere grafiske uttrykk som er utenfor kodifisert norm, men innenfor den operative.

Det er interessant å merke seg at av parene over, er <seil> og <segl>, <kull> og <kol>=/kå:l/ og <snø> og <sne> alle listet i Vanviks uttaleordbok, mens ingen ord på <snøgg>- er med.¹ At både <seil> og <segl> er med, er ikke så overraskende, i og med at begge er godkjente skrivemåter for et ord som er med i SØ, som er det Vanvik søker å beskrive i boken. At <snø> og <sne> er med, er heller ikke så overraskende, selv om <sne> ikke er del av den offisielle normen. Både /snø:/ og /sne:/ er del av SØ, som Vanvik skal beskrive, og han kunne vanskelig liste opp begge disse SØ-uttalene under oppslaget *snø*. <snø> kan ikke sies å representere /sne:/, slik jeg er inne på i diskusjonen om <trappa> og <trappen> over. At han ikke har noen oppslagsord på *snøgg*-, er heller ikke overraskende, i og med at vi hevder at ord på *snøgg*- ikke er del av SØ, som er hva Vanvik skal beskrive.

Hva så med *kol*?² Dette er mest sannsynlig et uttrykk for at Vanvik har et videre SØ-begrep enn min intuisjon tilsier. Han lister også former som *lauv*, *hue*, *trugsmål*, som jeg ville plassere utenfor SØ.

Eksemplene med *kol*, *snøgg* og *sne* over illustrerer en viktig egenskap ved bokmålsnormen. Den kodifiserte normen er ikke identisk med den faktiske normen som språksamfunnet faktisk forholder seg til; den faktiske normen er både videre og smalere enn den kodifiserte. Et søk i Oslo-korpuset på de alternative formene viser dette tydelig:

¹Bøyningsformer er ikke med i NUOV, så vi vet ikke om Vanvik anser begge formene *trappa* og *trappen* for å være del av SØ.

²*Kol* er også et synonym for *munda*, en etnisk gruppe i India, men det er lite trolig at det er denne betydningen Vanvik har hatt i tankene, selv om NUOV ikke oppgir ordforklaringer og slike usikkerhetsmomenter derfor kan oppstå.

<i>Søkekriterium</i>	<i>Antall treff</i>	<i>Søkekriterium</i>	<i>Antall treff</i>
kull	199	kol	0 ¹
hurtigtog*	9	snøggto*	0
hurtig*	1300	snøgg*	9
trappen	284	trappa	78
segl*	144	seil*	1032
løv	52	lauv	12
hode	326	hue	55
sne	86	snø	526

Asterisken indikerer at søket er utført med trunkerte søkekriterier, slik at søket skal treffe alle ord som *begynner* med de bokstavene som er angitt som søkekriterium. Jeg har brukt trunkerte søk der det ikke er fare for at dette skal medføre treff på andre morfem enn de aktuelle. Vi ser at det er ingen forekomster av *kol* eller *snøggto**, som ventet. De 9 forekomstene av det trunkerte søket på *snøgg* fordeler seg på 4 treff på idrettslaget *Snøgg* og 5 treff på former av *snøggbåt*, hvorav 2 skriver seg fra tekster på nynorsk. Det er interessant at dette ordet ser ut til å kunne brukes i bokmål, og det er sannsynligvis et kulturelt lån, der bokmål har lånt både begrepet og ordet fra "nynorsk kultur" på vestlandet. Verken *snøgg* eller *snøggto** ser imidlertid ut til å være del av den faktiske bokmålsnormen. Både *trappa*, *segl*, *lauv*, *sne* og *hue* er imidlertid i bruk; dog er de vesentlig mindre frekvente enn sine konkurrenter. *Hue* er imidlertid overrepresentert i de skjønnlitterære tekstene, og når vi avgrenser søket til avis- og ukebladstekster, ser vi at de 12 forekomstene utelukkende er direkte gjengivelse av tale eller del av det muntlige faste uttrykket "høl i hue".

Generelt er den faktiske bokmålsnormen slik den kommer til uttrykk i skriftlig språkbruk i offentligheten, en relativt smal norm i forhold til den kodifiserte normen. Men i enkelte tilfeller er det også slik at den faktiske normen inkluderer stavemåter som ikke er innenfor den kodifiserte normen; den første er med andre ord ikke en undermengde av den andre. I noen tilfeller sammenfaller formene

¹*Kol* gav faktisk tre treff i korpuset, men disse var alle referanser til Kolosserbrevet i bibelen.

som ligger utenfor bokmålsnormen med den uoffisielle riksmålsnormen, slik den er kodifisert av Riksmålsforbundet, for eksempel <sne>, <hverken>, men i enkelte andre tilfeller ligger den faktiske normen utenfor begge disse kodifiseringene, for eksempel <juggel>, <lompe>¹.

<i>Søkeord</i>	<i>Oslo</i>	<i>Usenet</i>	<i>Søkeord</i>	<i>Oslo</i>	<i>Usenet</i>
sne	86	273	snø	526	2771
hverken	895	10629	verken	1214	6511
sprog	52	1651	språk	1425	12046
jugl	1	0 ²	juggel	5	74
lumpe	0	0 ³	lompe	4	59

Et eksempel til fra Usenet-korpuset kan belyse to av poengene over. Søket er gjort på de bestemte formene av ordet for å unngå feiltreff på de homografe adjektivene.

<i>Søkekriterium</i>	<i>Antall treff</i>
hullet	4330
holet	1 ⁴
hollet	2
hølet	387

<Hullet> og <holet> er sidestilte former i den kodifiserte normen, mens <hollet> og <hølet> ikke er med. Tallene viser tydelig at den faktiske normen i realiteten inneholder bare <hull> og ikke <hol>. Faktisk er formen <hølet> mye mer

¹<lompe> erstattet nylig <lumpe> i den kodifiserte bokmålsnormen. Dette var en direkte følge av at <lompe> var så godt som enerådende i faktisk språkbruk, blant annet blant lompeprodusenter.

²2 treff på stedsnavnet <Jugl> er fjernet.

³8 metaspråklige treff, der endringen i stavemåte er diskutert, er fjernet.

⁴Antall treff på <holet> var faktisk 83, men 82 av disse forekomstene var i nynorsktekster, til tross for at søket var satt opp til bare å gjøres i bokmålstekster. Jeg vet ikke hvilke mekanismer som brukes for å skille mellom ulike språk eller målformer i Usenet; de fungerer i hvert fall ikke feilfritt. Søket på <hollet> er tatt med for å vise at mangelen på treff ikke skriver seg fra at det er brukt en annen bestemt form av *hol* enn den riktige, altså <holet>.

populær og prosentvis omtrent like populær som <segl> i forhold til <seil>, selv om <hølet> i motsetning til <segl> ikke er del av den kodifiserte normen.

Imidlertid indikerer ikke dette nødvendigvis at <hø> er del av den faktiske bokmålsnormen, og det lave antallet forekomster i Oslo-korpuset antyder at <hø> ikke er. Imidlertid illustrerer det det faktum at stilen på Usenet er så uformell og/eller muntlig at noen av språkbrukerne ikke oppfatter det som viktig å følge normen, verken den kodifiserte eller den faktiske, men heller bruker en staveform som i større grad reflekterer talemålsformen.

2.2.3 Bokmålets mønster

Formålet med oppgaven er å avdekke mønstre i bokmålsrettskrivningen og dens relasjoner med SØ talemål. I den sammenhengen må det gjøres avgrensninger som har med import å gjøre.

En tradisjonell dikotomi av et språks ordtilfang er mellom *arveord* og *lånord*, slik det omtales for eksempel i Lundeby 1987¹. Arveord er ifølge Lundeby "ord som alltid har vært i språket", mens Sandøy 2000 definerer arveord noe mer presist som "ord som har levd i språket så langt tilbake vi kan rekonstruere tidlegare språkstadium." Siden det ikke er snakk om lån i den ordinære betydningen av ordet, følger vi Sandøy og bruker benevnelsen *importord* om den andre gruppen av ord, altså de som vi kjenner til et opphav til i et annet språk. Sandøy deler så importordene inn i to grupper, som han kaller *fremmedord* og *lånord*, der *lånord* er de importord som er tilpasset det norske systemet fonologisk, morfologisk og ortografisk, og som man må ha fagkunnskap for å kunne skille fra arveordene. Dette er altså en smalere gruppe enn den som er omfattet av *lånord*-begrepet til Lundeby, mens *fremmedord* da blir importord som på minst ett av disse nivåene ikke er tilpasset norsk system, for eksempel:

- [tʃøtni], som strider mot norske fonotaktiske regler ved å ha sekvensen [tʃ] i fremlyd,

¹Artikkelen står i Mæhle, Lundeby og Grønvik 1987.

- [¹dʒåjstik], som strider mot det norske fonemsystemet med fonen [ʒ] eller [dʒ],
- *fan* – *fans*, som strider mot det norske bøyningssystemet for substantiv,
- <psyke>, som strider mot norske grafotaktiske regler ved å ha sekvensen <ps> frontalt,
- <tape>=/tæjp/, som strider mot norske grafonomiske regler ved å representere diftongen /æj/ med den diskontinuerlige grafemsekvensen <a>_<e>,
- <taxi>, som bruker grafemet <x>, som ikke er med i den grafemmengden som brukes til å representere norske ord med.

Jeg tror det i visse sammenhenger må være formålstjenlig også å inkludere et *semantisk* kriterium i skillet mellom lånord og fremmedord. Ord som er fullstendig tilpasset, eller som som en følge av tilfeldigheter eller delvise sammenfall i ulike språks strukturer følger de norske fonologiske, morfologiske og ortografiske systemene, kan likevel føles som fremmedord av språkbrukerne på grunn av et fremmedartet semantisk innhold, altså kulturelle forskjeller. Et eksempel er <samurai>=/samu¹raj/ – *samuraien* – *samuraier* – *samuraiene*, som er fullstendig tilpasset norsk på alle uttrykksplan,¹ men som innholdsmessig oppleves som fremmed på grunn av dets nære tilknytning til japansk kultur. Denne typen ord vedkommer likevel ikke diskusjonen her.

Sandøy 2000 slår sammen gruppene arveord og lånord og kaller unionen av disse for *hjemlige ord*. Hjemlige ord er altså alle ord som ter seg som norske, og det virker naturlig at det er disse ordene vi må studere for å avdekke mønstre i bokmålsrettskrivningen og dennes relasjoner til SØ talemål.

Men her oppstår et sirkelargumentasjonsproblem. Vi kan ikke avgrense gruppen av hjemlige ord ut fra kriterier som angår de ortografiske mønstre som

¹Dette forutsetter at (1) trykk på annet enn første stavelse og (2) diftongen /aj/ aksepteres som deler av norsk fonologi. Begge har sin opprinnelse i importord og er en indikasjon på import, men begge må også sies å være assimilert i språket og føles "helt norske" av dagens språkbrukere, jf. ordene *avis* og *kai*. Et eksempel som ikke ville være avhengig av disse forbeholdene, er <sumo>=/¹su:mo/.

kjennetegner bokmålssystemet, før vi har beskrevet disse mønstrene, og vi kan ikke beskrive de ortografiske mønstrene i bokmålssystemet før vi har skilt ut gruppen av hjemlige ord som eksempler på disse mønstrene. Det er ingen vitenskapelig metodisk løsning på dette, men vi kan se på mulige alternative kompromissløsninger:

1. Vi definerer bokmålsmønstrene ut fra gruppen av arveord alene.
2. Vi definerer lånord kun ut fra fonologiske og morfologiske kriterier.
3. Vi bruker vår intuisjon om hva som faller innenfor normmønstrene og definerer hjemlige ord ut fra disse mønstrene.
4. Vi bruker vår intuisjon som utgangspunkt i en iterativ prosess der vi gradvis finjusterer gresedragningen mellom hjemlige ord og fremmedord på grunnlag av en stadig mer nyansert beskrivelse av de ortografiske mønstrene.
5. Vi studerer lærerveiledninger eller andre arbeider innenfor grunnleggende skriveopplæring og bruker de mønstrene vi finner der, som definisjonsgrunnlag.
6. Vi definerer bokmålsmønstrene på grunnlag av unionen av hjemlige ord og fremmedord ut fra en argumentasjon om at bokmål og SØ utgjøres av hele ordtilfanget, og ikke bare de hjemlige ordene.

Alternativ 1 er problematisk ettersom vi i utgangspunktet ikke kjenner til eventuelle ulikheter mellom mønstrene som gjelder i hjemlige ord og naturaliserte importord. Sandøy nevner for eksempel at de fleste ord med /p/ i fremlyd (og <p> frontalt) er importord;¹ et bokmålsmonster basert på arveord alene vil derfor kunne inneholde en regel om at <p> ikke kan stå frontalt:

$$\neg \langle p \rangle | \# _$$

Dette er kontraintuitivt og neppe tjenlig ut fra formålet med beskrivelsen. Det er sannsynlig at det finnes mønstre fra én eller flere familier av naturaliserte

¹En del lydord er sannsynlige unntak.

importord som er tatt opp i bokmål, og som derfor utgjør en del av det vi ønsker å studere.

Alternativ 2 unngår sirkelargumentasjonen, men vil inkludere ord som de fonologisk og morfologisk fullstendig tilpassede <pizza>=/^lpitsa/, <sex>=/seks/ og <yoga>=/^ljå:ga/ og dermed fonem-grafemkorrespondanser som /ts/= <zz>, /ks/= <x>, /j/= <y> som vi intuitivt oppfatter som fremmede og som utvilsomt er sjeldne i bokmål. Imidlertid er ikke frekvens i seg selv et kriterium vi ønsker å bruke, fordi det vil utelukke mange arveord og potensielt mange skrivemønstre som en konsekvens. Vi kan heller ikke bruke det kjente fremmedspråklige opphavet til disse ordene, fordi dette igjen per definisjon vil utelukke alle importord, slik at dette i prinsippet vil være det samme som alternativ 1.

Alternativ 3 representerer på en måte en løsning der man innser at det er umulig å bryte sirkelargumentasjonen og tar tyren ved hornene. Det virker sannsynlig at en erfaren skriftspråkbruker har en sterk intuisjon om de mønstrene som styrer ortografien, og det er også et spørsmål om det ikke er nettopp denne intuisjonen vi ønsker å studere!

Alternativ 4 indikerer at det er formålstjenlig å justere hypotesen om mønstre underveis om man innser at den opprinnelige formuleringen av intuisjonen ikke var presis. Legg merke til at det er snakk om å revidere *formuleringen* av intuisjonen og ikke selve intuisjonen.

Alternativ 5 ville være interessant i et perspektiv der en granskning av lese- og skriveopplæring, pedagogikk eller lese- og skriveforskning stod i fokus, men det er lite trolig at praktiske lese- og skriveopplæringsverk har presise nok formuleringer til at disse kan brukes som utgangspunkt for å finne en beskrivelse av bokmålsmønstrene generelt.¹

Alternativ 6 medfører de samme problemene som er nevnt under alternativ 2, bare i større omfang; beskrivelsen ville nemlig også måtte omfatte grafem-representasjon av ord som ikke er fonologisk tilpasset, det vil si blant annet ord

¹ Dette har jeg imidlertid ikke gjort noe konkret forsøk på å finne ut.

som inneholder foner som ikke er en del av foneminventaret i SØ. Mange slike ord vil ha en uttale som er preget av kildespråkets fonetikk og fonologi, for eksempel <jingle>=[¹dʒɪŋɡl], <know-how>=?[¹nåohao]. Formulering av regler for hvordan bokmål representerer slike ord, vil raskt innebære formulering av ortografiske mønstre for samtlige språk som norsk har importert ord fra, og det blir både et for omfattende prosjekt og et prosjekt med et resultat som vil være av mindre interesse enn om vi greier å avgrense regelsettet til bokmålssystemet. Det vil dessuten her være en vanskelig grenseoppgang mellom hva som er importord som er tatt opp i språket med mangelfulle fonologiske tilpasninger, og hva som vil være kodeskifte.

2.3 Allomorfi

Diskusjonen omfatter ikke forholdet mellom allomorfer eller valg av allomorfer. Med andre ord er det ikke en diskusjon om "radikale" kontra "moderate" eller "konservative" former som <trappen> ~ <trappa> eller <kull> ~ <kol>, og det er heller ikke en diskusjon omkring selve systemet med en vid norm med mange valgmuligheter. Her vil vi bare se på forholdet mellom de skriftlige formene og de muntlige formene de kan sies å representere, altså <trappen>=/¹trapen/, <trappa>=/¹trapa/, og vi vil også utelate bokmålsformer som representerer uttaleformer som er utenfor standard østnorsk eller perifere med hensyn til denne varieteten. <kull>=/kul/ er dermed med i diskusjonen, mens <kol>=/kå:ɾ/, /kø:ɾ/ faller utenom.

2.4 Synkronisitet

Mange studier av rettskrivning har et diakront fokus, der talemålsutvikling, rettskrivningsreformer og enkeltords etymologi er utgangspunkt for diskusjonen.

Denne undersøkelsen er en synkron undersøkelse av dagens skriftsystem. Diskusjonen omfatter prinsipielt ikke etymologiske eller diakrone forhold, selv

om diakrone forhold åpenbart ligger til grunn for mange av de mønstrene vi skal diskutere.

Forholdet mellom hjemlige ord og fremmedord har likevel diakrone aspekter som vi har diskutert over og som nødvendigvis må spille en rolle i diskusjonen.

2.5 En lingvistisk undersøkelse

2.5.1 Tekstlingvistik

Diskusjonen omfatter ikke syntaktiske forhold eller forholdet mellom skrift og tale på det pragmatiske eller stilistiske nivå, slik dette blir behandlet for eksempel i Svennevig, Sandvik og Vagle 1995. Vi skal begrense diskusjonen til det fonologiske, det morfologiske og det leksikalske nivået; vi strekker oss utover ordet kun i de tilfeller der det leksikalske nivået kan sies å strekke seg utover ordet, som for eksempel i <i hvert fall>, <ta på>=/²ta:på/="ikle seg".

2.5.2 Psykologi

Det pågår mye forskning rundt det psykologiske perspektivet på skrift, altså innen lese- og skrivepsykologi. Mye av forskningen dreier seg om mer overordnede nivå, altså forståelse og meningskonstruksjon (se for eksempel Dysthe 1993 og Austad 1997b¹), men mange arbeider behandler også de tekniske aspektene ved lesing og skriving. Imidlertid er fokus i svært mange av disse arbeidene barn, andrespråkselever eller voksne med lese- og/eller skrivevansker.

Elvemo 1986 og Bråten 1990 er blant dem som behandler de kognitive prosessene bak lesing og skriving og disse prosessenes relasjoner til rettskrivningen. Dette er et interessant perspektiv, og det er å håpe at mer kunnskap omkring rettskrivningssystemer generelt kan ha positiv virkning på forskning rundt lesingens og skrivingens psykologi.

¹Artikkelen står i Austad 1997a.

Bråten 1994 behandler forholdet mellom lesing og skriving både på teknisk nivå og på meningsnivå. Han advarer (s.17) mot å trekke slutninger om lesing av norske tekster på bakgrunn av forskning på lesing av tekster i andre skript med andre typologiske egenskaper, særlig engelsk, men bruker likevel i fortsettelsen slik forskning i sin behandling av norsk uten eksplisitt å ta særlige forbehold.

Det er lite sammenlignende forskning av denne typen, med unntak av en del arbeid om andrespråkslesing og –skriving, og vi må være forsiktige med å overføre resultater fra ett skript til et annet med kanskje svært forskjellige egenskaper.

2.5.3 Pedagogikk

Vi kommer ikke inn på pedagogiske perspektiver eller diskusjonen rundt syntetiske eller analytiske leseopplæringsmetoder, eller kombinasjonsmetoder som LTG, se for eksempel artikler i Wold 1996 eller Wiggen 1992, som har studert barns stavefeil.

Også dysleksiforskning, se for eksempel Høien og Lundberg 1991, vil kunne dra nytte av økt kunnskap om rettskrivningssystemer, men heller ikke dette perspektivet berøres direkte i denne oppgaven.

Mye nyere forskning har fokus på lesing og skriving blant andrespråksbrukere og lese- og skriveopplæring av disse.

2.5.4 Normeringsteori

Dyvik 1993, Dyvik 2003 og Brunstad 2000¹ diskuterer teoretiske perspektiver rundt normer og normering som vi alt har vært inne på, spesielt med hensyn til forholdet mellom foreskrevne og operative normer. Dette har direkte relevans med hensyn til forholdet mellom bokmål og riksmål, og forholdet mellom den foreskrevne, vide bokmålsnormen og den operative, smalere normen.

¹Artikkelen står i Brodersen og Kinn 2000.

Vikør 1994 har et mer praktisk perspektiv på utfordringer ved språkplanlegging generelt og normering spesielt. Det er flere viktige normeringsmessige problemstillinger som faller utenfor denne oppgaven slik den er definert, og de viktigste er kanskje hensynet til talespråkets dialekter og hensynet til tradisjonen.

En skriftnorm for et nasjonalspråk er ment å fungere for hele nasjonen og alle innbyggernes dialekter, og dersom det finnes en offisiell normgivingsinstans, slik det gjør i Norge, må den derfor ta andre hensyn enn forholdet til bare én talemålsvarietet, slik jeg gjør i denne oppgaven. En bokmålsnorm som bare skulle ta hensyn til SØ, kunne for eksempel tenkes å normere /²faʃe/=<fasje>, noe som ville fungere svært dårlig for dialekter som ikke assimilerer /rs/ → /ʃ/ slik SØ gjør.

Tradisjon er også et viktig moment i normeringsspørsmål. Stabilitet i rettskrivningsnormen bidrar til å binde samfunnet sammen, i både et kortere og et lengre perspektiv. Det er viktig at alle levende generasjoner bruker skriftkoder som ikke er mer forskjellige enn at de er gjensidig forståelige. Det har også en verdi at tidligere generasjoners kulturelle ytringer er forståelige for ettertidens generasjoner.

I visse sammenhenger er det også formålstjenlig å opprettholde den fremmedspråklige kodifiseringen av importord for begreper fra andre kulturer, nettopp for å signalisere fremmedartethet. Dette kan i mange tilfeller også bidra til å motvirke homografi, som for eksempel i

<kick>=/kik/="metaforisk spark" : <kikk>=/çik/="titt"¹

¹I andre tilfeller er derimot virkningen motsatt, som i <tape>,<teip>=/tæjp/ : <tape>=/²ta:pe/.

3 Strukturen i standard østnorsk

Dette kapitlet gir en kort oversikt over det språkssystemet vi regner med i SØ. Mer detaljerte fremstillinger med diskusjon er integrert i kapitlene om grafonomien der dette er relevant.

3.1 Fonologi

3.1.1 Fonemsystemet

Når man skal diskutere det grafiske uttrykkets forhold til det fonemiske nivået, bør det ikke herske tvil om hvilken fonologiske beskrivelse som ligger til grunn. Vi bruker en strukturalistisk modell og følger Haslev 1985, Sivertsen 1967, Vanvik 1979 og deres beskrivelse av SØ fonemstruktur. Imidlertid er det slik at fonemet er en abstraksjon, og det er normalt ikke mulig å slå fast et språks fonemstruktur uten forbehold. Vi vil derfor her bare gi en oversikt og indikere områder som kan diskuteres. I diskusjonen om bokmål grafonomi vil vi så diskutere de ulike mulighetene og deres innvirkning på grafonomien.

SØ har ni vokalfonemer som alle kan opptre både korte og lange: /i y e ø æ a å o u/. Vi regner [ə] som en allofon av /e/. I minimale par som

<sparaes>=[²sparəs] : <spar-ess>=[²sparaes]

<urett>=[²u:ret] : <uret>,a=[²u:rət]

kan distinksjonen analyseres som en fonemisk distinksjon mellom segmentalene /e/ og /ə/, eller som en trykkdistinksjon mellom tilstedeværelse og fravær av bitrykk på den andre stavelsen med automatisk veksling av realiseringen av /e/ som [e] eller [ə] avhengig av bitrykk. Det er fordeler og ulemper forbundet ved begge analysene, men vi velger den siste, altså

<sparaes>=²spa:res/ : <spar-ess>=²spa:r.es/

<uret>=/²u:ret/ : <urett>=/²u:₁ret/.

Vi kan da si at bitrykket (delvis) realiseres ved allofonen [e] av /e/.

/æ/ har begrenset distribusjon og forekommer i hovedsak foran /r/ og retrofleksene inkludert [ɾ], som i <delje>=/²dæljɛ/=[dæɾjɛ], /²deljɛ/=[deljɛ]. Til tross for begrenset distribusjon er det liten tvil om /æ/'s fonemiske status.

Når det gjelder /u/, /o/, /å/, er det heller ingen tvil om deres fonemiske status. Det finnes minimale par for både lange og korte vokaler:

<Unn>=/un/ : <ond>=/on/ : <onn>, <ånd>=/ån/

<mur>=/mu:r/ : <mor>=/mo:r/ : <mår>=/mår:/

Historiske prosesser har imidlertid medført en del allomorfi med veksling mellom forskjellige par av disse vokalene:

<plomme>=/²plome/, /²plåme/

<bru>=/bru:/, <bro>=/bro:/

<hugg>=/hug/, <hogg>=/håg/

Alle disse variantene er del av den offisielle bokmålsnormen, og jeg tror også det kan argumenteres for at de er med i SØ. Som vi ser, blir noe av variasjonen reflektert i rettskrivningen og noe ikke.

SØ har 21 konsonantfonemer:

- 4 ustemte plosiver: /p t ɾ k/,
- 4 stemte plosiver: /b d ɖ g/,
- 4 nasaler: /m n ŋ ɳ/,
- 5 ustemte frikativer: /f s ʃ ç h/,.
- 2 stemte frikativer: /v j/,
- 2 likvider : /l r/.

Vi regner retrofleks lateral [ɭ] som en allofon av /l/. Sivertsen 1967 skrev allerede for snart 40 år siden at et sammenfall av de to enhetene, som tidligere har hatt klar fonemisk status, trolig er undervegs. Det er utvilsomt retrofleks uttale i mange ord som tidligere har hatt ikke-retrofleks uttale, etter alle vokaler bortsett fra /å a/¹, for eksempel <trolig>, <mulig>, <hele>. Det er vanskelig å finne minimale par i utlyd etter /a å/; en kandidat er /²å:ɭe/=<årle> : /²å:le/=<åle>, men *årle* er neppe del av SØ dagligtale. Et subminimalt par er [ja:ɭ]=<jarl> : [sa:l]=<sal>.² Retrofleks uttale blir i så fall en allofon av fonemet /l/, og det medfører en utvidet distribusjon av /æ/, som da kan opptre også foran /l/=[ɭ], som i /²hæ:li/=[²hæ:ɭi], og ikke bare foran /l/=[ɾ], som i /fæ:l/=[fæ:ɾ].

Vi regner også retrofleks flapp [ɾ] som en allofon av /l/. Vanvik 1979 og Kristoffersen 2000 skriver begge at flappen er et kjennetegn på mer folkelig uttale. Kristoffersen påpeker at den er mer vanlig postkonsonantisk enn postvokalisk, og postkonsonantisk er det ingen mulighet for minimale par, altså av typen [skli:] : [skɾi:]. I fremlyd alene kan den ikke forekomme, men i utlyd har den særlig utbredelse i negativt ladede ord, som *fæl*, *jåle*, *belje*. Både Kristoffersen og Vanvik understreker at SØ ikke kan ha [ɾ] i posisjoner med historisk *rð*, det vil si der normaluttalen er med /r/.

Ettersom Kristoffersen skriver innenfor et generativt paradigme og dermed benytter et annet fonembegrep enn meg, og Vanvik ikke eksplisitt nevner om han omtaler fonemikk eller fonetikk, er det vanskelig å konkludere absolutt omkring flappens fonemiske status ut fra denne litteraturen. Et mulig minimalt par er [ku:l]=<cool> : [ku:ɾ]=<kul>, men det er tvilsomt om [ɾ] i *kul* kan sies å tilhøre

¹Bak /æ/ har ikke SØ hatt /l/, med unntak av ved allofonen [ɾ], som omtalt over.

²Retrofleks uttale er vanlig hos dagens barn og unge også etter /a å/, som i <ball>=[baɭ], og dette tyder på at sammenfallsprosessen har fortsatt og kan nærme seg fullbyrdelse. Mangelen på minimale par er i hvert fall ingen bremse på denne prosessen.

SØ. Vi kaller [ɾ] en allofon av /l/, og dermed kan også kort /æ/ sies å kunne opptre foran /l/, som i <belje>=²bælje/>[2bærje].

Dette blir til sammen 21 konsonanter, og totalt 30 segmentale fonemer.

SØ har videre tre prosodemer, trykk, kvantitet og tonem, som opptrer i delvis solidaritet. I hvert ord er det én stavelse som er potensiell bærer av trykket ved aksentuering. Denne stavelsen har lengde, som knyttes enten til vokalen eller til den etterfølgende konsonanten eller konsonantsekvensen. En del stavelser er også overlange, men den distinktive funksjonen er knyttet til *vokalens* lengde. De overlange stavelsene er med få unntak bøyde eller avledede former, for eksempel <bilt>=/bi:lt/="av bile", <svinsk>=/svi:nsk/="av svin", der rotmorfemets vokalkvantitet er overført til den bøyde eller avledede formen. Imidlertid finnes også røtter med overlange stavelser, for eksempel <påske>=²på:ske/>[KV:K:V]. Dersom den trykksterke stavelsen er etterfulgt av en trykksvak stavelse, har den trykksterke stavelsen et distinktivt tonem, tonem 1 eller tonem 2. Som nevnt over tilegner vi også bitrykk fonemisk status på grunnlag av distinktiv funksjon i visse ord.

Antall stavelser har også potensielt betydningsskillende funksjon, for eksempel i <rottene>=²råtne/>[S-S-S] : <råtne>=²råtne/>[S-S]. Segmentalt og med hensyn til de tre suprasegmentale trekkene som er beskrevet over, er disse ordene identiske. Distinksjonen er knyttet til antall stavelser, der den ekstra stavelsen i det første ordet er realisert med en konsonant som stavelseskjeerne, en stavelsesbærende /n/.

Også plassering av stavelsesgrense kan i visse tilfeller være distinktivt, for eksempel i et minimalt par som

<u-tøy>=²u:-tøj/>[²u:t^høj] : <ut-øy>=²u:t-øj/>[²utøj].

Distinksjonen stavelsesgrense er i dette tilfellet realisert som aspirasjon eller fravær av aspirasjon etter /t/. Aspirasjon har normalt ikke distinktiv funksjon i

norsk, men er en automatisk veksling som er avhengig av fonologiske omgivelser, f.eks. <kake>=/²ka:ke/=[²k^ha:ke].

3.1.2 Fonotaks

Jeg har allerede vært kort inne på begrensninger i distribusjonen av enkelte fonemer, nemlig /æ/ og prosodemene trykk, kvantitet og tonem. Jeg skal bare kort påpeke enkelte fonotaktiske regler som er relevante for diskusjonen av grafonomien under.

/æ/ har en sterkt avgrenset distribusjon i standard østnorsk. Den forekommer bare foran /r/ og retrofleksene, inkludert allofonene [ɾ] og [ɻ] av /l/.

- <der>=/dæ:r/, <merr>=/mæ:r/
- <lært>=/læ:t/, <ert>=/æɾ/
- <ferdig>=²fæ:ɖi/, <verdi>=/væ¹ɖi:/
- <herlig>=²hæ:li/=[hæ:ɻi], <fæl>=/fæ:l/=[fæ:ɾ], <delje>¹=/²dælje/=[dærje]
- <særs>=/sæ:ʃ/, <vers>=/væʃ/

Når det gjelder stavelsesstrukturen i SØ, så består hver stavelse av en obligatorisk kjerne, som kan ha en konsonant eller en konsonantsekvens foran seg og en konsonant eller en konsonantsekvens etter seg. Kjernen består av en vokal eller en stavelsesbærende konsonant, en nasal eller en likvid. Hvilke sekvenser som kan stå foran og etter kjernen, styres av forskjellige regler, for eksempel kan vi ha tre konsonanter foran kjernen bare dersom den første er /s/. Utlydsreglene er ulike avhengig av om stavelsen utgjør ett morfem eller flere.

Visse konsonantfonemer har begrenset distribusjon også når de opptrer alene, altså ikke som del av en konsonantsekvens:

/h/ opptrer kun frontalt.

¹ Uttale også /²delje/ ifølge NUOB og NUOV. NROB har ikke noe om uttale av verbet, men skriver "helst" /²dærjer/ om substantivet *deljer*.

- /ç/ opptrer kun frontalt, med unntak av intervokalisk i de to leksemene *bikkje* og *krykkje*="fugl".
- /j/ opptrer kun frontalt. Det betyr at vi analyserer ord som *bejaende*, *lojal* og *rujern* med stavelsesgrense foran /j/. I ord som *kajakk* og *jojo* analyseres [j] som del av en diftong.
- /ŋ/ opptrer kun stavelsesfinalt¹ og kun etter korte vokaler.
- /t/ /d/ /n/ opptrer kun finalt i isolerte stavelser. I sammenhengende ytringer vil ytre sandhi kunne frembringe retroflekse konsonanter ordfrontalt i sammenhenger som *har du*, *har tatt*, *har ny*. /d/ og /n/ opptrer bare etter lange vokaler, mens /t/ kan opptre også etter korte: <sart>=/sa:t/, <fart>=/fa:t/.
- [l] har de samme begrensningene som de andre stemte retrofleksene.²

3.2 Morfologi og morfofonemikk

Jeg vil anse hovedtrekkene i morfologien i SØ som kjent.

Når det gjelder morfofonemikken, skal jeg bare nevne kort visse egenskaper av relevans for diskusjonen av grafonomien under (Haslev 1972) .

Det er få – om noen – regelmessige vokalvekslinger i SØ, slik det finnes i språk som engelsk, tysk og russisk. Om jeg hadde definert [ə] som et eget fonem, og ikke en allofon av /e/, kunne vi kanskje operert med en automatisk veksling mellom /e/ og /ə/ basert på trykk, men ellers er det ingen automatiske vokalvekslinger. De i historiske framstillinger mye omtalte avlyds- og omlydsvekslinger³ er ikke synkront regelmessige.

Det er imidlertid noen automatiske vekslinger i andre områder:

¹Intervokaliske forekomster tolkes som stavelsesfinale, som i <finger>=/¹fɪŋ-er/, <hangar>=/¹haŋ-ga:r/.

²[l] opptrer riktignok også som assimilasjonsprodukt i fremlyd for eksempel etter /ʃ/.

³For eksempel i Torp og Vikør 2003.

Veksling i stemthet. Stemte konsonanter mister gjerne sin stemthet i assimilasjon med ustemte plosiver. Dette gjelder oftest i omgivelser med /t/, ettersom /t/ opptrer frekvent som bøyningsmorfem. Eksempler:

- <stiv>=/sti:v/ ~ <stivt>=/sti:ft/
- <legge>=/²lege/ ~ <lagt>=/lakt/

Veksling mellom plosiv og null. I noen typer ord er det veksling mellom ustemte plosiver og null. Eksempler er:

- <huset>=/¹hu:se/ ~ <husets>=/¹hu:sets/
- <vennlig>=/²venli/ ~ <vennligst>=/¹venlikst/

Veksling i kvantitet. Utlydsvokaler blir gjerne korte i bøyde omgivelser, for eksempel:

- <grå>=/grå:/ ~ <grått>=/gråt/
- <gå>=/gå:/ ~ <gått>=/gåt/

I moderne SØ er det imidlertid ingen automatisk veksling i kvantitet avhengig av antall konsonantlyder i utlyd. De lange vokalene beholdes oftest også foran konsonantsekvenser, for eksempel:

- <fin>=/fi:n/ ~ <fint>=/fi:nt/
- <bile>_{,v}=/2bi:le/ ~ <bilt>=/bi:lt/

Men:

- <løpe>=/2lø:pe/ ~ <løpt>=/løpt/
- <leke>=/2le:ke/ ~ <lekt>=/le:kt/, /lekt/

Eksempelene viser at her er det variasjon.

4 Grunnleggende skriftteori

Det er ikke denne oppgavens formål å formulere en allmenngyldig teori om skrift, og ei heller å tilby et resymé og en kritikk av alle tidligere arbeider innen området skriftteori. Oppgavens formål er å studere forholdet mellom bokmål og standard østnorsk, og jeg vil avgrense de teoretiske betraktningene til hva som er formålstjenlig ut fra oppgavens problemstilling. Dette betyr ikke at vi akter å unngå teoretiske problemstillinger, men jeg ønsker ikke å bruke tid og plass på å lage et teoretisk rammeverk som nødvendigvis er treffende for alle typer skrift som har eksistert gjennom alle tider; jeg ønsker kun å lage en teoretisk grunnmur som kan bære det (sørøst-)norske hus.

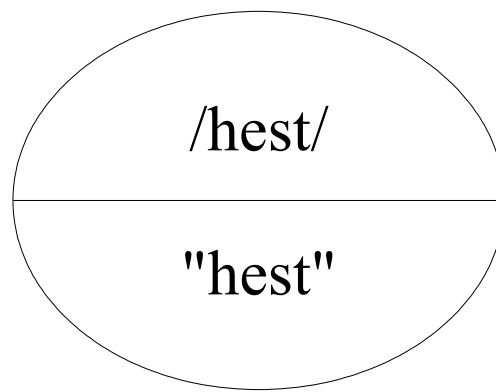
Det kan oppstå dilemmaer når man prøver å konstruere en strengt teoretisk ramme for en vitenskapelig behandling av et fenomen som vi allerede har et godt utbygd, skjønt noe mangelfullt, dagligspråk om. Man kan velge å kaste vrak på alle termer og konsepter fra dagligspråket, eller man kan søke å inkorporere dagligspråkstermene i det vitenskapelige rammeverket. Fordi dagligspråkstermene er så godt kjent og godt innarbeidet, og mye brukt også i vitenskapelige sammenhenger, og fordi de representerer fruktbare innsnevringar av de nye termene vi trenger, er det den siste framgangsmåten vi velger her.¹ Det har også den fordelen at det lar oss diskutere problemstillinger med allmennspråkstermene før vi har utviklet en vitenskapelig terminologi.

¹En tredje mulighet ville være å bruke termene fra dagligspråket, men gi dem nytt innhold. Denne framgangsmåten har den ulempen at man i så fall til enhver tid må ha på det rene om teksten er fagspråklig eller allmennspråklig, og det er neppe hensiktsmessig for leseren å la samme termen ha to ulike innhold avhengig av register. Det ville også øke belastningen på forfatteren.

4.1 Skrift og språk

4.1.1 Definisjon av skrift

Det er ingen enighet i fagmiljøet om hva skrift er og hvordan det kan defineres. Vi tar utgangspunkt i en strukturalistisk modell og Saussures tegnbegrep (Saussure [1915]:65)¹ og prøver å innpasse en teori om skrift i teorien om talespråk som tegnsystem. Saussures språktegnbegrep er en tilfeldig forbindelse mellom fonologisk uttrykk (/hest/) og semantisk innhold ("hest"):



Saussure sier (s. 23) at "Språk og skrift er to distinkte tegnsystemer; det andre har som eneste formål å representere det første." Sampson 1985:26 definerer skrift som et system for å representere ytringer i et talespråk ved hjelp av varige, synlige merker.² Papazian 1984:40, derimot, mener at "skriftspråket er ikke bare funksjonelt likeverdig med talemålet som uttrykkssystem; det er et eget, selvstendig tegnsystem, som følger sine egne normer [...]."

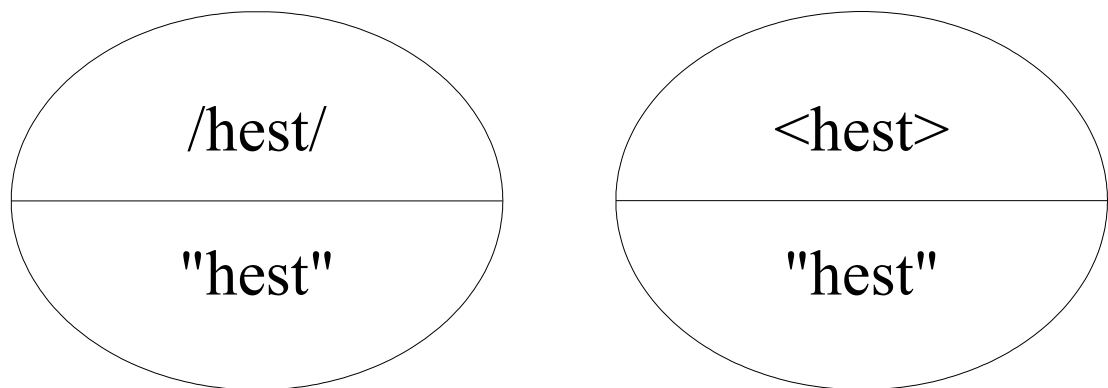
Jeg vil for dette formålet bruke en definisjon av skrift som baserer seg på skrift som en representasjon av talt språk; jeg følger altså Saussure og Sampson og ikke Papazian. Dette er ikke ment fullstendig å underkjenne definisjoner av skrift som et separat tegnsystem basert på skriftspråkets særegne stilistiske trekk, men den valgte definisjonen er en definisjon som for vårt formål fungerer godt.

¹Henvisninger til sidetall i Saussure [1915] refererer til sidetall i den engelske utgaven som er gitt i litteraturlista.

² Min oversettelse. Den engelske originalen lyder "a system for representing utterances of a spoken language by means of permanent, visible marks."

Det kan tenkes minst tre ulike framstillinger av skriftens forhold til talen:

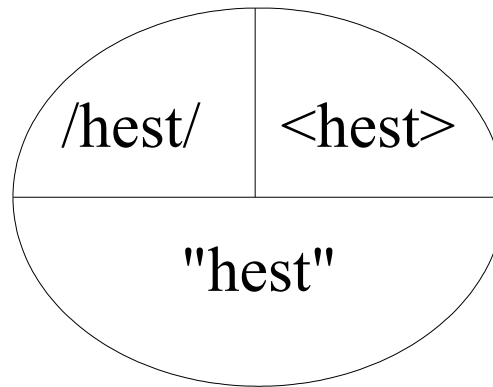
For det første kan skrift ses på som et separat tegnsystem uavhengig av tale, slik blant andre Papazian 1984 gjør.



Dette er en definisjon som gjør det vanskelig å diskutere skriftens formelle forhold til talen videre, slik jeg ønsker å gjøre i denne oppgaven. Hva er fonemografisk skrift dersom skrift er et tegnsystem som ikke har noen forbindelse til talen? Hva er det seksåringer gjør når de "knekker lesekode"? Hva innebærer det å normere et skriftspråk eller å fastsette et normgrunnlag for det (Vikør 1994)? Hva vil det si å lese høyt eller å skrive ned hva noen sier?

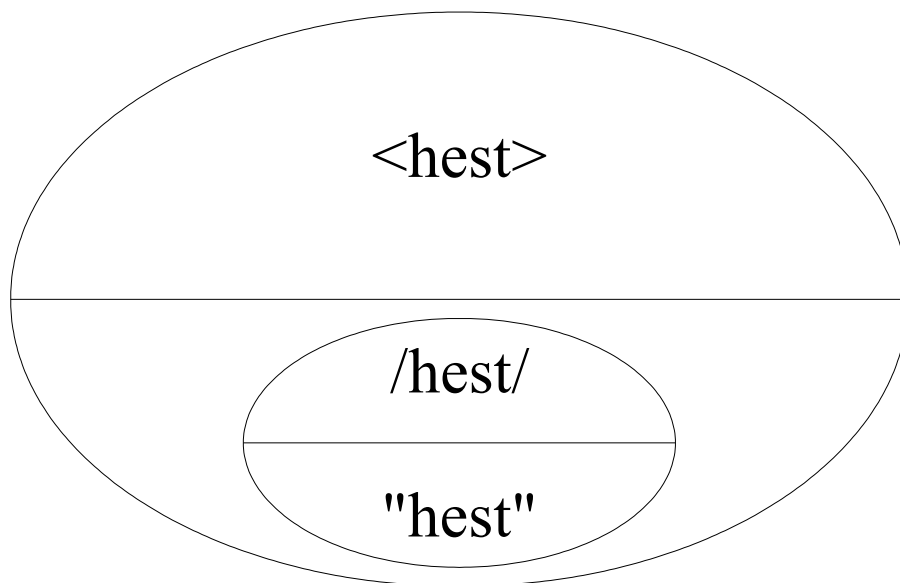
Papazian innrømmer riktignok at de to tegnsystemene har mange fellestrekk, og at de påvirkes av hverandre, men det er vanskelig å forklare hvordan de semantiske strukturene i skrift og tale skal kunne være uavhengige av hverandre. I prinsippet vil da `<hest>` og `/hest/` kunne ha helt ulikt semantisk innhold, for eksempel `"ku"` og `"hest"`, eller det ene uttrykket kan ha et videre innhold enn det andre, for eksempel `"firbent husdyr"`. Vi kjenner imidlertid ikke til noen skrifttradisjon der den semantiske strukturen i henholdsvis skrift og tale er prinsipielt forskjellige. Det er i høyden snakk om stilistiske forskjeller, selv om ulikhetene i arabisk skrift og tale kan være store (Sampson 1985).

For det andre kan skrift ses på som et tegnsystem parallelt med talesystemet, altså som et eget uttrykk for det samme språksystemet.



Her er det semantiske systemet felles, mens uttrykkssystemene er prinsipielt urelaterte. Dette forklarer hvorfor vi kan være ganske sikre på at innholdet forbundet med par av fonologiske og grafonomiske uttrykk er det samme, men det forklarer ikke de indre strukturelle likhetene mellom de to uttrykksystemene. Det blir fortsatt vanskelig å forklare hvorfor forholdet mellom antall fonemer og antall grafemer i et uttrykkspaar som oftest er ganske nær en-til-en i alfabetisk skrift, og hvordan vi kan sette opp regler som styrer de fleste unntakene. Det er fortsatt vanskelig å forklare begrepene fonemografisk eller syllabisk skrift og å beskrive hva seksåringene gjør når de "knekker lesekode". Vi kan kanskje forklare fenomenene "høytlesning" og "notering av tale" med denne modellen, men det blir vanskelig å forklare den systematiske sammenheng mellom synonymer og deres respektive stavemåter eller mellom allomorfer i de to uttrykksystemene. Hvorfor sier vi /sne:/ når vi leser <sne> og /snø:/ når vi leser <snø>? Hvorfor sier vi /^ltrape/ om vi leser ordet <trappen> selv om vi vanligvis sier /^ltrapa/ når vi ikke leser?

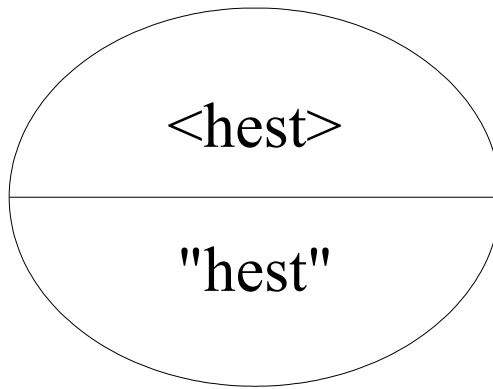
For det tredje kan skrift ses på som en representasjon av talt språk – i motsetning til av tale:



Dette betyr at skriften representerer *hele* språktegnet, og ikke bare uttrykket eller bare innholdet (Saussure [1915]:23). Det betyr at det finnes en systematisk og samtidig relasjon mellom både grafonomisk uttrykk og fonologisk uttrykk på den ene siden, og grafonomisk uttrykk og semantisk innhold på den andre, og modellen lar oss studere de spørsmålene vi har stilt over. Derfor er dette den modellen vi har valgt, og vi kaller det utvidede tegnet i figuren over for det *grafonomiske tegnet*.

Den fjerde muligheten er en naiv oppfatning av skrift som muligens er utbredt blant legfolk, nemlig at det er en representasjon av tale. Barn som nettopp har knekket lesekoden, vil sikkert også ha sympati for en slik modell, for det grunnleggende prinsippet for alfabetskrift¹ er nettopp det at én bokstav representerer én språklyd, og det er forståelsen for dette prinsippet barnet har tilegnet seg når vi sier at det har knekket lesekoden og kan skrive *godt* som <GÅT>.

¹Termen er her brukt preteoretisk. En mer presis terminologi vil bli utviklet etter hvert.



En slik definisjon støtter seg på egenskapene ved grafemet i motsetning til egenskapene ved det grafonomiske språktegnet, og det er problematisk å bruke en slik definisjon blant annet fordi det blir vanskelig å forklare heterografe representasjoner av homofoner, morfemografisk skrift, og andre iboende egenskaper i de fleste eksisterende skriftsystem.

Lydskrift, for eksempel fonemisk transskripsjon, er riktignok en representasjon av tale, selv om lydskrift også er en konvensjon og ofte også representerer ikke-fonologiske strukturer, som for eksempel ordgrense. Dessuten er det mange aspekter ved talen som lydskrift gjerne ikke representerer. Basert på Saussures definisjon av skrift over, er imidlertid ikke lydskrift skrift, i og med at formålet med lydskrift ikke er å representere *tegnene* i talen, men bare det fonologiske uttrykket. Lydskrift som fenomen vil ikke bli behandlet i denne oppgaven.

Heller ikke tegnspråk kan sies å være skrift, og ikke bare på grunn av dets manglende bruk av varige merker på et medium. Tegnspråk representerer ikke talespråk. Tegnspråk *er* et eget tegnsystem – altså et språk – som eksisterer parallelt med talt språk i et samfunn, og selv om det vil være parallelle strukturer i tegnspråk og talt språk i et samfunn, slik det vil oppstå parallelliteter i ethvert tospråklig samfunn, er det ingen prinsipielle eller formelle forbindelser mellom SØ eller andre dialekter av norsk på den ene siden og norsk tegnspråk på den andre.

4.1.2 Semasiografi

Semasiografiske systemer er ifølge Sampson 1985 grafiske symbolsystemer som ikke er skrift fordi de ikke representerer ytringer i talt språk. Eksempler er matematisk notasjon og noter for musikk. Begge disse notasjonssystemene representerer *kunnskap* ved hjelp av et system av varige, synlige merker, slik en mer vidtfavnende definisjon av skrift kunne lyde. Harris 1995 definerer skrift på denne måten og inkluderer både matematisk notasjon og musikknotasjon i mengden av skriftsystemer, men ingen av disse symbolsystemene representerer ytringer i talt språk, så de faller utenfor min definisjon av skrift.

Musikknotasjon representerer ikke noe som kan gjengis ved tale i det hele tatt, bortsett fra ved en indirekte og langtekkelig gjengivelse, som for eksempel kunne lyde som:

Én stav. C-dur. Fire fjerdedelstakt. Fjerdedelsnote enstrøken C, fjerdedelsnote enstrøken D, fjerdedelsnote enstrøken E, fjerdedelsnote enstrøken F, to halvnoter G, fire fjerdedelsnoter A, helnote G...

for de fire første taktene av "Lisa gikk til skolen". Denne ytringen ville med ganske stor sannsynlighet gjøre en notekyndig i stand til å produsere den intenderte grafikken, skjønt det er flere elementer av underforstått informasjon her, blant annet hvilken nøkkel det skal skrives i, hvor taktstrekene skal stå, og at også G-ene og A-en skal være enstrøkne. Også antall kryss eller b-er må deduseres ut fra den informasjonen som er gitt og allmenn musikkunnskap. Imidlertid er det åpenbart at den musikkgrafikken som produseres, ikke *representerer* den talte ytringen over, men derimot en sekvens av toner i tillegg til strukturelle egenskaper som takt og toneart.

Lesing av matematisk notasjon er vanligere, og matematiske formler kan også dikteres, men det er heller ingen en-til-en-relasjon mellom matematiske symboler og noen lingvistisk enhet på noe nivå. Matematisk notasjon kan derfor ikke være et tegnsystem med språk tegn som innhold.

Et så enkelt matematisk uttrykk som $10 - (2 \times 3) = 4$ kan leses på mange ulike måter, for eksempel som

1. *Ti minus to ganger tre er lik fire.*
2. *På den ene siden av likhetstegnet har vi fire, mens på den andre har vi differensen mellom ti og to ganger tre.*
3. *To ganger tre fra ti blir fire.*
4. *Produktet av to og tre subtrahert fra ti er fire.*
5. *Ti minus parentes begynnelse to kryssmultiplisert med tre parentes slutt er lik fire.*

Ingen av disse formuleringene kan sies å være mer korrekte enn andre som talte representasjoner av uttrykket, men de har noe ulik status. Alternativ 1 er potensielt tvetydig, selv om det grafiske uttrykket ikke er det. Betydningen av dette alternativet kan nemlig påvirkes ved pauser i lesningen etter henholdsvis *ti* og *to*. Matematikere vil imidlertid oppfatte det som entydig ved en nøytral lesning, ettersom det er en iboende egenskap ved det matematiske symbolsystemet at multiplikasjon har høyere prioritet enn addisjon og subtraksjon, slik at parentesene i uttrykket strengt tatt er overflødige.

Alternativ 2 til 4 gjengir alle de relevante matematiske egenskapene som er gjengitt av grafikken, men ingen av dem kan garanteres å produsere den samme grafikken ved å dikteres til en matematiker. Alle de følgende uttrykkene – og mange flere – er *synonyme* med uttrykket over med hensyn til det matematiske innholdet:

$$\begin{aligned}4 &= 10 - (2 \times 3) \\10 - 2 \times 3 &= 4 \\-(2 \times 3) + 10 &= 4\end{aligned}$$

og de kunne alle tenkes å bli produsert etter ytringene 2 – 4 over. Kun alternativ 5 vil med særlig stor sikkerhet produsere det opprinnelige uttrykket, og det er fordi dette er en lesning som i stor grad har karakter av å være en *transskripsjon*, altså en gjengivelse av hvert symbol i rekkefølge, nesten som lydskrift. Med mer

komplekse formler, blir avstanden mellom notasjon og talespråk enda mer markert og muligheten for å "lese" uttrykkene høyt på denne måten mindre. Akkurat som for musikkseksemplet over er det klart at matematikkgrafikken ikke kan sies å gjengi ytringer, men matematisk informasjon.

Semasiografiske systemer er derfor ikke skriftsystemer i egentlig forstand. Resten av kapitlet vil dreie seg om teori om og typologisering av egentlige skriftsystemer. Vi kaller egentlige skriftsystemer for *skript*.

4.2 Grafemet

For å kunne typologisere skriftsystemer må vi introdusere grunnleggende teori om grafemet. Teorien om grafemet kaller jeg grafemikk, og den diskuteres i mer detalj i kapittel 4.4. Jeg bygger grafemikken på definisjonen av skrift i kapittel 4.1.1.

Når en lingvistisk teori om skrift skal funderes på Saussures tegnbegrep, slik jeg gjør det her, kan man fort komme inn i et terminologisk uføre i og med at ordet *tegn* allerede har en godt innarbeidet betydning i allmennspråket. Saussures tegn er en tosidig enhet av uttrykk og innhold, mens det allmennspråklige *tegn* gjerne brukes om bare det skriftlige uttrykket, eller en del av det, altså *symbolet*¹: bokstaver, skilletegn eller diakritika. Det kan være på sin plass å presisere de ulike begrepene og deres termer nå, selv om de presise definisjonene skal diskuteres mer inngående senere:

- *språktegn* eller bare *tegn* brukes om Saussures tegnbegrep som en tosidig forbindelse mellom et fonologisk uttrykk og et semantisk innhold.
- *grafonomisk tegn* brukes om det utvidede tegnbegrepet som ble omtalt i 4.1.1, og som danner grunnlaget for vår definisjon av skrift, der uttrykket er det grafiske uttrykket og innholdet er hele språktegnet.

¹Denne bruken av termen *symbol* strider mot Saussures terminologi, ettersom han mener at et *symbol* aldri er fullstendig arbitrært.

- *grafemisk tegn* brukes om en tosidig forbindelse mellom en grafisk figur, altså grafemet, og et fonologisk uttrykk, som oftest en figur. Dette begrepet blir nærmere omtalt i kapittel 4.2.3.
- *skriftegn* brukes i sin allmennspråklige betydning om en liten, grafisk uttrykksenhet. Som oftest vil ordet være synonymt med *grafem*, for eksempel <a>, <:>, <^>.

4.2.1 En gyldig lingvistisk enhet?

Det er vanlig å operere med begrepet *grafem* innenfor det vitenskapelige studiet av tekst, grafonomien.¹ Inspirasjonen er åpenbar; termen er formet som en parallell til begrepet *fonem* innenfor fonologien. Tilsvarende brukes gjerne termen *graf* som en parallell til *fon* og *allograf* som en parallell til *allofon*. Termenes likhet er ment å reflektere begrepenes likhet, og visse likhetstrekk i egenskaper er også åpenbare.

Om vi bruker Saussure [1915] som utgangspunkt, er fonemets sentrale egenskap funksjonell; det er den minste betydningsskillende språklyden. Som Jakobson og Halle 1956 påpeker, er det imidlertid mellom fonemer og grafemer noen vesentlige forskjeller som gjør at vi må opptre varsomt i omgang med grafem-begrepet. Hver grafisk enhet signaliserer en spesifikk enhet. Den representerer ikke nødvendigvis den samme enhet i alle omgivelser, men det finnes et samband mellom den grafiske enhet og en enhet i talespråket: et distinktivt trekk, et fonem, et morfem eller en del av et morfem, en stavelse, en semantisk enhet, en syntaktisk enhet, et prosodisk uttrykk, eller lignende. Det finnes derimot ikke noen forbindelse mellom den fonologiske enhet fonemet og noen annen enhet verken i et annet system eller i det fonologiske systemet. Fonemet signaliserer kun *forskjell*, eller *otherness* i Jakobson og Halles terminologi. Det vil si at fonemet er en *figur* (Saussure [1915]), mens grafemet i en viss forstand er et *tegn*, i og med at det har en uttrykksside, nemlig den grafiske form, og en innholdsside, nemlig hva det representerer – i fonemografiske skriftsystem ofte et fonem.

¹Allén 1971 o.a.

4.2.2 Grafemet som funksjonell enhet

Vi definerer her de grunnleggende termene.

En *graf* er et skrifttegn, altså et permanent, synlig merke brukt til representere en del av en ytring i et talt språk. Et *grafem* er det minste innholdsskillende skrifttegn. Legg merke til at dette ikke er helt parallelt med definisjonen av fonemet, som er den minste *betydningsskillende* språklyd. Imidlertid er innholdssiden av det grafonomiske tegnet hele språktegnet, slik at grafemets funksjon er å skille mellom ulike språktegn. Denne distinksjonen kan finne sted på tre ulike måter:

- på innholdssiden av språktegnet
- på uttrykksiden av språktegnet
- på begge sider samtidig

På begge sider samtidig er den vanligste situasjonen. Vi ser på eksempler for hver av situasjonene:

1. <sne>=/sne:/="frossent regn" : <snø>=/snø:/="frossent regn"

Dette viser at distinksjonen <e>:<ø> har innholdsskillende funksjon ved at uttrykksiden av språktegnene er ulike i de to grafonomiske tegnene.¹

2. <toll>=/tål/="avgift" : <tolv>=/tål/="12"

Dette viser at distinksjonen <l>:<v> har innholdsskillende funksjon ved at innholdssiden av språktegnene er ulike i de to grafonomiske tegnene.

3. <arm>=/arm/="legemsdel" : <orm>=/årm/="slange"

Dette viser at distinksjonen <a>:<å> har innholdsskillende funksjon ved at både innholdssiden og uttrykksiden av språktegnene er ulike i de to grafonomiske tegnene.

¹Dette er strengt tatt gyldig bare om vi først har vist at /e/ og /ø/ har betydningsskillende funksjon og dermed er ulike fonemer. Vi regner dette som u diskutabelt og markerer derfor språklydene med fonemklammer.

Lydmassen i tale er kontinuerlig både paradigmatiske og syntagmatiske. Det samme er tilfellet for håndskrift, og i tillegg til kommutasjonstester av typen over er det nødvendig å finne minimale par der grafemene kommuterer med null for å være sikker på at grafene faktisk er de *minste* språktegnskillende enhetene:

1. <bord>=/bo:r/="møbel" : <bor>=/bo:r/="har tilhold"
2. <kelnere>=/^lkelnere/="servitører" : <kelnerer>=/^lkelnerer/="servitører"
3. <øk>=/ø:k/="gammel hest" : <øks>=/øks/="huggeredskap"

I trykt skrift er i stor grad denne jobben alt gjort for oss, selv om det paradigmatiske og syntagmatiske diskrete i trykte typer ikke er noen garanti for de funksjonelle egenskapene. Vi må segmentere grafene både paradigmatiske og syntagmatiske på grunnlag av funksjon og ikke fysiske egenskaper. Vi skal ikke gjøre den detaljerte analysen for alle bokstavene i alfabetet her; for de minst frekvente bokstavene i alfabetet kan det være en utfordring å finne alle de nødvendige minimale parene, og vi vil bare gi eksempler her. Resten overlates til leseren som en øvelse.

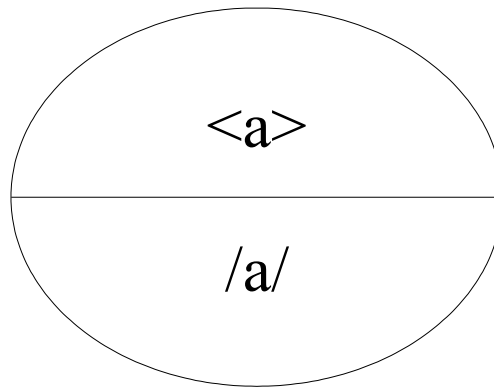
<celle> : <selle> : <elle>
<qua>¹ : <kua>
<watt> : <katt> : <att>
<sex> : <sel> : <se>
<zen> : <sen> : <en>

4.2.3 Grafemet som tegn

Vi kan også framstille forholdet mellom grafem og fonem, eller bokstav og lyd i et alfabetisk skript², som en slags tegnforbindelse, men der innholdssiden er fonemet, og uttrykkssiden er grafemet. Vi kan si at grafemet uttrykker fonemet, innholdet i tegnet, og kaller denne forbindelsen det *grafemiske tegnet*:

¹For <q> er det en ytterligere utfordring at bokstaven har så begrenset distribusjon; den forekommer nesten utelukkende foran <u>.

²Vi bruker ordet *alfabetisk* som en preteoretisk term her.



Dette er ikke helt parallelt med Saussures språk tegn, men om vi legger vekt på for det første den konvensjonelle forbindelsen mellom sidene, og for det andre arbitrariteten i forbindelsen, ser vi at forholdet mellom grafem og fonem har mange av språk tegnets egenskaper.

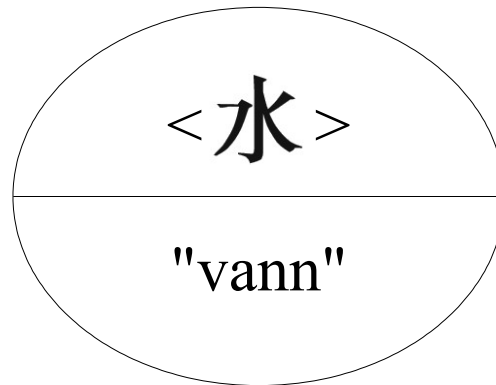
Det er en *konvensjon* i vårt språksamfunn – nærmere bestemt vårt skriftspråksamfunn – at fonemet /a/ vanligvis uttrykkes i skrift ved grafemet <a>. At det er en konvensjon, vil si at det er en allmenn overenskomst, og at det ikke er gitt noe enkeltmedlem av språksamfunnet den autoritet at han eller hun kan endre denne forbindelsen. Om departementet utstedte et dekret om at det norske skriftspråksamfunn fra nå av skulle representere fonemet /a/ med <w>, <α>, <水> eller <𐌆>, ville de fleste av oss likevel bruke <a> i morgen og sannsynligvis også om hundre år, fordi dette ville sikre at kommunikasjonen gled best.

Forbindelsen er *arbitrær* fordi det ikke er noe ved utseende til bokstaven <a> som gjør at det er mer naturlig at akkurat den formen representerer dette fonemet. Fonemet /a/ kunne teoretisk sett like gjerne vært representert ved <w>, <α>, <水> eller <𐌆>, skjønt i praksis vil det stilles visse krav til økonomi i både resepsjon og produksjon av grafemet.

Parallellen med språk tegnet svikter for det første ved at innholdssiden i forbindelsen kun er en figur og uten semantisk betydning, og for det andre ved at strukturen i innholdssiden ikke er tilfeldig; den er bundet av fonemstrukturen i språket. Det er altså ikke slik at hele det grafemiske tegnet er bestemt av

skriftspråkskonvensjonen alene slik at skriftspråkskonvensjonen er autonom; innholdssiden er bestemt og strukturert av talespråkskonvensjonen.

Det er verdt å merke seg at om vi prøver å framstille et grafemisk tegn for et skript som kinesisk, ville det være fristende å vise innholdssiden som et semantisk innhold:



I dette tilfellet er forbindelsen et ekte saussuresk tegn. Grafemet er her ikke en figur, og vi bruker ikke begrepet *grafemisk tegn*, men reserverer dét til tilfeller der begge sidene i forbindelsen er figurer. Ettersom grafemet ikke er en figur, inngår det i et grafonomiske tegn, med hele språktegnet på innholdssiden. Heller ikke i kinesisk er skriften knyttet direkte til innhold.

4.2.4 Allografi

Grafemet er som nevnt en funksjonell enhet. Parallelt med begrepet *allofoni* fra fonologien bruker vi termen *allografi* om det fenomen at grafer med ulike fysiske egenskaper kan være medlem av den samme funksjonelle enheten grafemet.

Ulike grafer i samme grafem kalles *allografer*, og eksempler på dette er

<a> ~ <a>

<g> ~ <g>

(Legg merke til at klammene her ikke markerer *grafemisk* transskripsjon, men kun at det er det grafiske uttrykket vi diskuterer. Vi har ikke reservert egne

transskripsjonsklammer for dette formålet ettersom det sjelden er aktuelt å gjøre denne distinksjonen. Se 1.2.1.) Selv om de fysiske formene har ganske store forskjeller, kan vi se at disse grafene tilhører samme funksjonelle enhet ved å undersøke om det finnes minimale par:

- <ba> : <ba>, <ak> : <ak>, <bak> : <bak>, <tatt> : <tatt>, <bag> : <bag>
- <gi> : <gi>, <god> : <god>, <og> : <og>, <vegg> : <vegg>, <seg> : <seg>

Ikke i noen posisjoner er det mulig å finne kommutasjon mellom disse grafparene; i alle disse parene er hele språktegnet som grafsekvensene representerer, det samme i begge tilfeller. I eksemplene med <a> og <g> er de fysiske forskjellene ganske store, men for de fleste andre bokstaver er de fysiske forskjellene mellom ulike allografer i trykt skrift relativt små, for eksempel:

- <b b b b b>
- <f f f f f>
- <p p p p p>
- <s s s s s>
- <t t t t t>

I håndskrift, derimot, kan forskjellene i fysisk utseende mellom ulike allografer være ganske påfallende:

- <b b>
- <f f>
- <p p p>¹
- <s s>
- <t t>

¹Den siste allografen av <p> her er vanlig å bruke på de britiske øyer, men den fysiske formen er kanskje så ulik prototypen til en <p> i bokmål at den ikke vil fungere som en allograf av <p> i bokmål.

I de fleste tilfellene er det snakk om fakultativ variasjon, men visse allografer kan også stå i mer eller mindre komplementær distribusjon. Særlig i håndskrift kan de grafiske omgivelsene være med å bestemme en grafs fysiske utseende. Dette er mindre vanlig i trykt skrift, men se likevel om ligaturer i 4.5.1.

Versaler og minuskler er ikke allografer, ettersom de kan danne minimale par:

<Per>=/pe:r/="mannsnavn" : <per>=/pær/="for hver"

4.2.5 Distinktive grafiske trekk?

I fonologien kan fonemene defineres som klynger av distinktive fonetiske trekk. For eksempel kan /p/ defineres som et fonem med trekkene {labial}, {plosiv}, {ustemt} og skilles fra /b/ ved trekket \pm {stemthet}, fra /t/ ved trekket {labial} kontra {dental}, og fra /m/ ved trekket {plosiv} kontra {nasal}, mens de to andre trekkene i hvert tilfelle er identiske.¹ Tilsvarende kan vokalene skilles ved å bruke et system av trekk som uttrykker henholdsvis åpningsgrad, artikulasjonssted og lepperunding. Ved å bruke distinktive trekk oppnår man en mer økonomisk beskrivelse av fonemsystemet ved at antall trekk som kan bestemme alle fonemer i et språk, vil være mindre enn antall fonemer i språket.

Om man som en begynnelse prøvde å ekstrahere distinktive grafiske trekk fra grafemmengden av versalene <A B C D E F G H I J K L M N O P Q R S T U V W X Y Z Æ Ø Å>, ville man raskt finne at regelmessigheten og dermed den potensielle økonomien i dette systemet ikke er av en slik art som i fonemsystemet. Teoretisk kan 29 fysiske enheter skilles ved minimum 5 binære egenskaper, men i praksis vil mange naturlige systemer ikke være så regelmessige at det er mulig å bruke det teoretisk laveste antall trekk. Haslev 1985 definerer for eksempel konsonantsystemet i bergensk med henholdsvis et binært system av seks trekk og et ikke-binært system av ti trekk, mens konsonantsystemets 17 fonemer teoretisk kunne bestemmes av fem binære trekk.

¹Jakobson og Halle 1956 videreutviklet dette systemet til et system av binære trekk, slik jeg har indikert for trekket {stemthet}. Det er ikke nødvendig å gå inn i dette for å illustrere denne teorien relatert til grafonomi, og det er også tvilsomt om fonemsystemets natur er binært i SØ (Haslev 1985).

Noen trekk vil være felles for mange av versalene og således potensielt redusere kompleksiteten i en beskrivelse av systemet:

- {loddrett strek} – <B D E F H I K L M N P R T Æ>
- {vannrett strek oppe} – <E F T Æ>
- {vannrett strek midten} – <A E F G H Æ Å>
- {vannrett strek nede} – <E L Z Æ>
- {"mage"} – <B C D G P R>
- {skråstrek ned mot høyre} – <A K M N R V W X Å>
- {skråstrek opp mot høyre} – <A K M V W X Z Æ Å>

Det er allerede nå lett å ane noen av problemene man kommer opp i:

1. Noen av trekkene er ikke primitive. Det ville ha vært ønskelig å kunne uttrykke slektskapet mellom {skråstrek oppover} og {skråstrek nedover}, men dette blir vanskelig i et binært system. I et ikke-binært system kunne man dele inn grafemene i tre grupper: de med skråstrek oppover, de med nedover og de med begge. Men da har vi allerede brukt tre trekk.¹
2. Noen av trekkene gjelder flere ganger for samme grafem, f.eks. {loddrett strek} for <M N H> og skråstrektrekkene for <W>.
3. Noen av trekkene trenger ytterligere kvalifiseringer for å definere egenskapen presist nok, for eksempel lengden av skråstrekkene eller "magens" orientering og størrelse.
4. De fleste trekkene trenger presis spesifisering av kontaktpunktet med andre trekk i grafemet, for eksempel skillet mellom <A V X>.
5. Mange av grafemene har helt særegne egenskaper som de er alene eller så godt som alene om, for eksempel {ring over} for <Å>, {liten hale} for <Q> og {slangeform} for <S>; <G J U Y> er rett og slett problematiske å definere på denne måten i det hele tatt uten å bruke relativt ordrike beskrivelser.

¹Det er også mulig å konstruere et trekkssystem i flere nivåer, som også ville kunne uttrykke dette ved tre trekk: $\pm\{\text{skråstrek}\}=(\pm\{\text{oppover}\}, \pm\{\text{nedover}\})$.

Av dette kan vi se at versalgrafemenes fysiske egenskaper neppe er regelmessige nok til at de kan beskrives av et system av distinktive trekk som er mer økonomisk enn en opplisting. For minusklene er dette kanskje enda verre.

Vi ser også at det ikke er noen sammenheng mellom grafemenes distinktive trekk og de distinktive trekkene til de fonemene de representerer. Vi kan ikke en gang se av grafemenes form om de representerer vokaler eller konsonanter – eller om de er semasiografiske grafemer for den del. Dette ser vi tydelig om vi ser på at i det greske alfabetet er <P>=[r], mens i cherokee representerer <D>=[a], mens <W> har det syllabiske innholdet [la] og <4> også er et fonografisk grafem med en fonologisk figur som innhold.¹ Dette illustrerer det arbitrære i det grafemiske tegn.

I fonemsystemet er enkelte av trekkene gjerne nøytralisert i visse fonem. For eksempel utgjør det ingen forskjell om en /l/ er stemt eller ikke, eller om den har lepperunding eller ikke. Det kan være relativt stor variasjon både i de fysiologiske og de akustiske variablene mellom ulike foner som likevel tilhører samme fonem, og som kommunikasjonsmottagere er vi generelt svært tolerante overfor de akustiske signalene vi mottar. I grafemikken er dette mer problematisk. Dersom et grafem låner et ekstra trekk som ikke er med i definisjonen av grafemet, vil resultatet som regel bli en graf som vanskelig vil kunne bestemmes. For eksempel vil <Δ Φ Γ Λ Π Σ Ξ> av bokmålsbrukere ikke kunne plasseres blant bokmålsgrafemene, selv om de er bestemt av de samme distinktive trekkene som vi har brukt over. Og likevel har vi sett at i håndskrift har vi ganske stor toleranse overfor allografer som ikke en gang har de samme distinktive trekkene, som for eksempel de ulike realisasjonene av <f> over.

Av dette går det fram at grafemer er ikke bare funksjon; de er også form.

¹Katzner 1986.

4.3 Skriftypologi

Sampson 1985 deler alle egentlige skriftsystem inn i to grupper: systemer der grafemenes primære tilknytning til språktegnet er via det semantiske innholdet, og systemer der grafemenes primære tilknytning til språktegnet er gjennom det fonologiske uttrykket. Den førstnevnte typen kaller han *logografi*, mens den sistnevnte typen omtaler han som *fonografi*. I og med at jeg har definert det grafonomiske tegnet som et tegn med hele talespråkstegnet som innhold, er en slik dikotomi problematisk. Skillet må baseres på hvorvidt grafemet er en figur eller ikke.

4.3.1 Morfemografi

Logografiske skriftsystemer er skriftsystemer der grafemene prinsipielt representerer et semantisk innhold, altså der det grafemiske tegnets innhold er semantisk. Fonografiske skriftsystemer er skriftsystemer der grafemene prinsipielt representerer en fonologisk figur, altså der det grafemiske tegnets innhold er fonologisk.

Bokmålsgrafemene <a>, <i>, <o> og <å> representerer vokalfonemene /a/, /i/, /o/ og /å/ og inngår normalt sekvensielle forbindelser med andre grafemer for å representere semantisk innhold, men disse vokalsgrafemene kan også representere semantisk innhold alene:

<a>=/a:/="preposisjon"

<i>=/i:/="preposisjon"

<o>=/o:/="tiltaleform"

<å>=/å:/="elv"

Dette gjør imidlertid ikke bokmål til et logografisk skript. For det første er det klart at dette er en perifer egenskap ved bokmålet, men viktigere er det at forbindelsen mellom figuren og det semantiske innholdet er i *språktegnet* og ikke i det *grafemiske* tegnet. Det er forbindelsen i språktegnet mellom fonemet /i/ og

innholdet "i" som gjør at grafemet <i> kan representere et semantisk innhold på egen hånd, og ikke en iboende egenskap ved grafemet <i>.

I praksis finnes det ikke skriftsystem som bruker enheter på noe høyere semantisk nivå enn morfemet, og logografiske skriftsystem kan derfor entydig omtales som *morfemografiske*. Foruten en del historiske, nå utdødde, systemer, er kinesisk det best kjente eksemplet på denne gruppen. Det er viktig å merke seg at også kinesisk er et ekte skriftsystem, altså et tegnsystem der innholdet i det grafonomiske tegnet er et talespråkstegn og ikke kun et semantisk innhold. Dette står i motsetning til den populære oppfatning at kinesisk er et slags universelt system for logisk notasjon, uavhengig av talen, slik Leibniz med sitt *characteristica universalis*¹ satte seg fore å lage.

DeFrancis 1989 og Coulmas 2003 er uenige i Sampsons klassifisering av kinesisk og kaller det et morfem-syllabisk system. De hevder i motsetning til ham at det ikke eksisterer semantisk baserte skriftsystemer. De mener at notasjonssystemer må utnytte den økonomiske fordel som språkets dobbelte artikulasjon tilbyr, for å kunne kalles skrift, og at det er praktisk og prinsipielt umulig å tenke seg skriftsystemer som ikke gjør dette. Dette er jo en interessant diskusjon, men for vårt formål er dette punktet ikke viktig, og vi måtte ha gått mye dypere i den kinesiske skriftens natur for å kunne diskutere dette.

4.3.2 Fonografi

Den andre hovedgrenen av skriftsystemer har grafemiske tegn der innholdet knyttes primært til det fonologiske uttrykket. Disse systemene kalles fonografiske, og de kan deles inn i de to hovedgruppene syllabografi og fonemografi. I syllabiske systemer er innholdssiden i det grafemiske tegnet en stavelse; hvert grafem representerer² altså en stavelse. Et eksempel på dette er japansk kana-skrift. Kapittel 6.1 omtaler syllabografi i mer detalj.

¹Omtalt i DeFrancis 1989.

²Termen *representere* defineres under.

I den andre typen av fonografiske systemer, fonemografi, representerer hvert grafem et fonem.¹ Det er vanlig å hevde at alle vesteuropeiske skriftsystem prinsipielt er av denne typen, selv om de, som vi skal se, kan ha trekk fra andre typer av systemer. Noen begreper og termer som angår fonemografi, kan introduseres her.

Når et grafemisk tegn inngår i en ordform, sier vi at tegnets grafem *representerer* tegnets fonem. Dette er egentlig bare en kortere måte å si at grafemet er uttrykket i det grafemiske tegnet som har fonemet som innhold. Det er viktig å merke seg at termen "representerer" skjuler det tosidige, arbitrære forholdet som ligger i tegnbegrepet, og i større grad indikerer en ensidig relasjon "står for" mellom grafem og fonem. Jeg understreker at jeg holder fast på den tosidige tegnrelasjonen, og at termen bare er en hendig uttrykksmåte. Vi utvider virkeområdet til termen og sier også at et grafem eller en grafemsekvens *representerer* andre språklige enheter, for eksempel fonemsekvenser i det grafemiske tegnet og morfemer i det grafonomiske tegnet. For eksempel representerer grafemsekvensen <str> fonemsekvensen /str/, og grafemsekvensen <kj> representerer fonemet /ç/, mens grafemet <t> kan representere nullmorfemet {bestemt form nøytrum} i et ord som <eplet>=²eple/="frukten".

Motsatt sier vi at fonemet *lisensierer* grafemet og utvider også denne termens virkeområde tilsvarende.

Vi definerer tentativt et grafems *primærfonem* som det fonem som grafemet mest frekvent eller i de minst markerte omgivelsene representerer. Mer generelt kan vi snakke om et grafems eller en grafemsekvens' *primære fonologiske innhold* der innholdet består i noe annet enn ett fonem.

¹Det er dette som gjerne blir omtalt som alfabetskrift. Dette er imidlertid en problematisk term i og med at den også kan ha andre innhold. I allmennspråket snakker man gjerne om "det kinesiske alfabetet", som inneholder grafemer som *ikke* har fonemografiske egenskaper, og i matematikken betyr alfabet rett og slett domenemengde av symboler. I den matematiske forstand vil alfabetet av bokmålsgrafemer bestå av alle skilletegn, diakritika og tall i tillegg til bokstavene. (Se for eksempel Partee, ter Meulen og Wall 1993.)

Tilsvarende definerer vi et fonems *primærgrafem* som det grafem som fonemet mest frekvent eller i de minst markerte omgivelsene lisensierer. Mer generelt kan vi snakke om et fonems eller en fonemsekvens' *primære grafiske uttrykk* der uttrykket består i noe annet enn ett grafem.

Et viktig prinsipp i fonemografiske systemer er *linearitetsprinsippet*. Dette prinsippet sier at rekkefølgen av grafemene i en ordform er direkte diktert av rekkefølgen av fonemene de representerer. Linearitetsprinsippet er også relevant der grafemene representerer andre typer enheter enn fonemer.

Fonemografiske systemer blir behandlet mer detaljert i kapittel 4.5.

4.3.3 Trekkbasert skrift?

Ikke noe kjent skriftsystem kan sies systematisk å representere allofonisk variasjon, men koreansk står likevel i en særstilling ved systematisk å representere distinktive trekk i fonemene. Hvert skriftsymbol i koreansk er satt sammen av et visst antall grafiske elementer, der hvert element representerer ett fonetisk trekk i fonemet. De grafiske elementene er dessuten ikke arbitrære, men ikoniske, i det de stilistisk gjenspeiler tungens posisjon ved artikulasjon av de ulike fonemene.

Vi regner de latinske bokstavene for fullstendig arbitrære, selv om de utvilsomt har en fortid som motiverte symboler. Forløperne til våre bokstaver var stiliserte tegninger av objekter som hadde navn som hadde grafemets primærfonem i fremlyd, for eksempel <D>, fra gresk <Δ>=*delta*, som opprinnelig representerte en dør (sannsynligvis i et telt). Mange av de opprinnelige betydningene kan lett gjenkjennes i de nåværende latinske bokstavene om man kjenner dem; for eksempel var <A> en tegning av en okse, der bokstavens "ben" var oksens horn, og trekanten oksens hode. (Det er lettere å se om man snur bokstaven på hodet eller sidelengs.)

Om man leter etter ikonisitet i det latinske alfabetet i dag, kan man kanskje tenke seg at en <O> er motivert ut fra leppenes runding ved uttale av denne lyden.

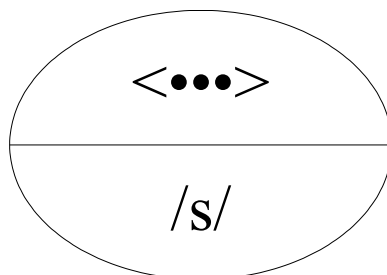
Dette er sannsynligvis en tilfeldighet, men vi kan finne tendenser til sammenhenger mellom grafiske trekk i bokstavene og fonologiske trekk i fonemene de representerer.

1. Alle vokaler utenom <y> har x-høyde; det vil si at de har bokstavenes minimale høyde.¹ Mange av vokalene (<e o u æ ø>, og i hvert fall allografer av <a å>) har en form som er dominert av en kurve.
2. Alle plosiver har høyde som er lengre enn x-høyde; de går enten over (<b t d k f>) eller under (<g>) midtfeltet.

For begge punktene over gjelder at også andre bokstaver har de aktuelle trekkene, men disse egenskapene representerer like fullt tendenser, og det er sannsynlig at gode lesere ubevisst gjør bruk av kunnskap om slike tendenser (Hellmark 2000).

4.3.4 Metaskrift

Analogt med at metaspråk er språk om språk, kan vi snakke om metaskrift som skrift som representerer skrift. Samfunnet har en rekke spesialiserte systemer som brukes når det av utenforliggende årsaker er vanskelig å bruke vanlig skrift, for eksempel morse, staving i tegnspråk, semafor, ASCII. Disse systemene representerer bare et annet skriftlig uttrykk; de inngår i en tegnforbindelse med grafemene i det ordinære skriftsystemet på innholdssiden:



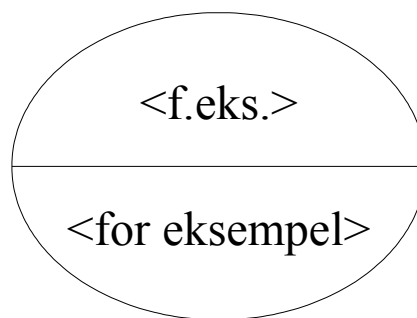
¹I engelsk er det vanlig å klassifisere <y> som en konsonant, selv om den brukes til å representere både vokaler og konsonanter. Om man følger denne konvensjonen, gjelder vokalens x-høyde unntaksløst for det engelske skriftsystem.

og det er utelukkende skriftens grafonomiske egenskaper som gjør at vi kan lese disse metaskriftlige systemene:²

<— ——— •••>=<mos>=/mo:s/="knust masse"

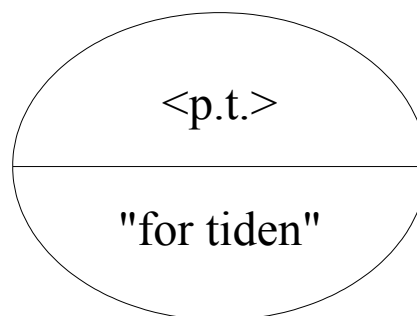
Mer interessant i denne sammenhengen er diverse forkortelsessystemer som er i allmenn bruk i bokmål og andre fonemografiske skript. Det er minst tre ulike måter slike forkortelser kan forholde seg til skriftsystemet på:

1. Rent metaskriftlig, som i <f.eks.>=<for eksempel>=/får ek¹sempøl/



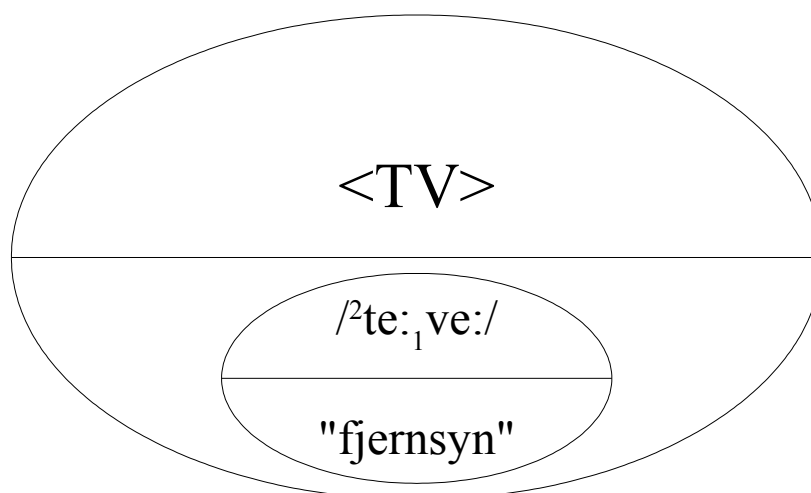
Her representerer rett og slett forkortelsen et annet skriftlig uttrykk.

2. Morfemografisk, som i <p.t.>=/¹pe:¹te:/="for tiden". Synkront er det neppe mange språkbrukere som kjenner til forbindelsen <p.t.>=<pro tempore>, så <p.t.> fungerer uavhengig av dette. Forbindelsen er sterkere direkte til innholdet, og mange vil nok lese <p.t.> som /få l¹tj:dn/.



²Figurene i morsesystemet er bare tre: prikk, strek og mellomrom, og systemets dobbelte artikulasjon må sies å være mellom disse figurene og de kombinasjonene som de kan inngå i, og som kan utgjøre innholdssiden i morsens tegnbegrep som i figuren.

3. Som akronymer som får leksemisk status, både i skriftspråket og i talespråket. Et eksempel er <TV>=²te:₁ve:/,²te:ve/. Det fonologiske uttrykket er her motivert av bokstavnavnene. Formene *televisjon* eller *televisjonsapparat*, som er forkortelsens etymologiske bakgrunn, spiller mindre rolle i det synkrone systemet.



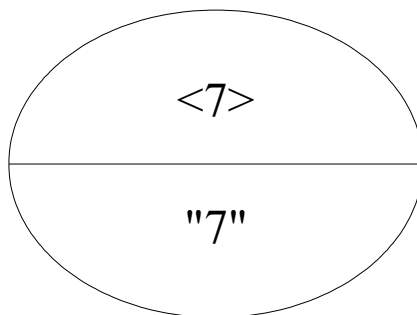
En annen form for forhold mellom grafisk notasjon og talt språk har vi der vi bruker tallgrafemer i fonemografisk skrift, som i

<Guri har 7 epler.>
<Guri har 17 epler.>

I det første tilfellet kan det se ut som om vi har et nytt tilfelle av morfemografisk skrift, der <7> representerer morfeminnholdet "7" direkte uten å gå veien om det fonologiske uttrykket:

<7>="7"=/sy:v/, /ʃu:/

Særlig i og med at vi har to alternative fonologiske uttrykk for denne betydningen i SØ, altså to allomorfer, virker dette som en god modell:



Imidlertid støter vi raskt på problemer med denne modellen for det andre eksemplet. Hvis $\langle 1 \rangle = \{\text{én}\}$ og $\langle 7 \rangle = \{\text{sju}\}$, burde $\langle 17 \rangle^* = \{\text{én sju}\}$ ifølge linearitetsprinsippet, men vi har jo at $\langle 17 \rangle = \text{"syttten"} = \text{²søtṇ/}$, som enten må analyseres som en rot eller antagelig som en sammensetning eller avledning med *ten* som sisteledd eller suffiks. Hvis $\langle 17 \rangle$ hadde vært et uanalyserbart grafem, hadde ikke dette vært noe problem, men $\langle 17 \rangle$ må analyseres som to grafemer $\langle 1 \rangle + \langle 7 \rangle$, og disse grafemene forholder seg ikke systematisk eller lineært til noen talespråklige enheter, verken på fonologisk eller morfologisk nivå. Enda verre blir det hvis Guri hadde 2707 epler. Hvilken fonologisk eller morfologisk enhet representerer $\langle 0 \rangle$ her?

Vi har allerede sett at matematisk notasjon tilhører den gruppen av notasjonssystemer som vi kaller semasiografiske, og eksemplene over viser at de skrifttegnene vi kaller tall, og som vi bruker i løpende tekst i vårt fonemografiske system, beholder sine semasiografiske egenskaper når de blir brukt innimellom fonemografiske skrifttegn.

4.4 Grafemmengde og alfabet

4.4.1 Bokstaver og alfabet

En undermengde av grafemene i bokmål er de grafemene vi preteoretisk kaller *bokstaver*. Det er allment akseptert at det norske alfabetet har 29 bokstaver.

Dette er noe som er så innprentet i oss at det kan være vanskelig å fri seg fra en

slik oppfatning, men det er en kjensgjerning at vi har dobbelt sett: minuskler og versaler¹.

Etter en lingvistisk analyse er det mye som taler for at vi er best tjent med å operere med en grafemmengde bestående av 58 bokstavgrafemer:

- *Betydningsskillende funksjon.* Vi har minimale par som <hans> : <Hans> og <sola> : <Sola>. I det siste tilfellet er det distinksjon mellom både betydningsinnhold og uttaleuttrykk², mens i det første tilfellet er det distinksjon mellom betydningen i to homofoner.
- *Liten grad av trekklikhet.* Noen par har stor grad av likhet, f.eks. <o> : <O>, som bare skilles ved relativ størrelse, mens andre par har svært liten grad av likhet, f.eks. <a> : <A>, eller <l> : <L>, der minuskelen ligner mest på en annen bokstavs versal <I>.

På den annen side har vi at følgende holder:

- Samme primærfonem.³
- Komplementær distribusjon i mange tilfeller, f.eks. etter punktum.
- Meningsfylte tekster kan skrives med bare versaler og likevel oppfattes som normerte. Tekster bestående av bare minuskler har litt mindre akseptert status, selv om dét også forekommer, særlig i uformelle elektroniske medier

¹ Jeg bruker terminologien fra Hellmark 2000 her. Termen majuskel for versal er også i bruk.

² Normalt ville vi her ha snakket bare om innhold og uttrykk, men med definisjonen av det grafonomiske tegnet blir begrepene innhold og uttrykk tvetydige. "Uttrykk" kan betegne både fonologisk uttrykk og grafisk uttrykk, mens "innhold" kan betegne språklig innhold, altså betydning, og det *skriftlige tegnets* innhold, altså språktegnets to sider.

³ Et mulig unntak til dette er {x, X}. <x> har primært fonologisk uttrykk /ks/ i bokmål, som i <sex>=/seks/ og <taxi>=/tækxi/, mens <X> kanskje må sies å ha primært fonologisk uttrykk /s/ i og med at <X> bare forekommer i fremlyd og dermed må uttales uten plosiv foran sibilanten, for eksempel <Xantippe>=/san2tipe/, <Xerxes>=/særkses/ og <Xylofon>|<.>_=/sylo¹fo:n/. Imidlertid bør dette tilskrives fonotaktiske regler i SØ og ikke egenskaper ved grafemikken eller grafonomien i bokmål.

som Usenett og SMS, og neppe virker særlig hemmende på kommunikasjonen.

- Historisk likt opphav. Dette er selvfølgelig ikke noe argument i seg selv, da det finnes både fonempar (for eksempel /k/ og /ç/) og grafempar (for eksempel <u> og <v>) med felles opphav. Men det felles opphavet har jo medført at for alle bokstavene har versalen og minuskelen samme navn, og vi bør heller ikke se bort fra den sterke tradisjonen språksamfunnet vårt har for å definere alfabetet vårt som bestående av 29 bokstaver.

Vi sier at bokmål har 29 bokstaver. Hver av bokstavene kan opptre som versal eller som minuskel. En bokstav er dermed et ordnet par $\{a_{\text{minuskel}}, a_{\text{versal}}\}$. Legg merke til at vi må definere bokstaven som et *ordnet* par i og med at det ikke er tilfeldig hvilket medlem som er minuskel og hvilket som er versal. Det ligger imidlertid ikke noen betydning av sortering eller prioritet i denne ordningen, selv om vi vet at versalen kom først rent historisk (Sampson 1985).

Mengden av bokstaver kaller vi alfabet. Alfabetet er dermed en delmengde av grafeminventaret og består av disse 29 ordnede parene, og dette stemmer godt med den preteoretiske oppfatningen vi har av alfabetet.

Bokstavene har to egenskaper som er særegne for disse grafemene:

1. De fleste bokstavene har et utvetydig fonologisk *bokstavnavn*.
2. De fleste bokstavene har et *primærfonem*, en *basisuttale*.

Bokstavnavn og primærfonem er prinsipielt urelatert, men i praksis beslektet.

Når man benevner en bokstav i tale, er det vanlig å benytte bokstavnavnet, mens det er svært uvanlig å benytte et grafisk uttrykk for dette navnet i skrift.¹

Eksempler på bokstavnavnene er $\{a, A\} = /a:/$, $\{b, B\} = /be:/$, $\{c, C\} = /se:/$, $\{j, J\}$

¹ Det er mye som taler for at dette er prinsipielt umulig. Bokstaver som skal omtales i skrift, betegnes rett og slett ved ett av bokstavens grafiske uttrykk. Et mulig unntak er en metaspråklig omtale av bokstavene der det fokuseres nettopp på bokstavenes navn. All omtale av bokstaver er, nærmest per definisjon, metaspråklig, men bare en fonologisk metaspråklighet som er rettet spesifikt mot bokstavnavnet, kan tenkes å bruke et skriftlig uttrykk som <be>, <ce>, <ku>, <dobbeltve>.

=/je:/, /jåd/, {q,Q}=/ku:/, {w,W}=/^ldåbelt,ve:/. Bokstavnavnet er en del av skriftspråksamfunnets tradisjon. Normalt denoterer bokstavnavnet hele det ordnede paret, men hvert medlem kan ved behov omtales spesielt ved å kvalifisere bokstavnavnet med enten *stor* eller *liten*. Med unntak av for {j, J} er bokstavnavnene utvetydige for bokmål, og de må defineres ved opplisting.

Det fonologiske primæruttrykket er et mindre entydig begrep, selv om også denne størrelsen har en forankring i språksamfunnets tradisjon. Et forsøk på definisjon kan være det fonemet som grafemet mest frekvent representerer, eller det fonemet som mest frekvent lisensierer grafemet, noe som ikke nødvendigvis gir samme resultat. Et problem med en slik kvantitativ definisjon er at det er vanskelig å bestemme en entydig metode for slik kvantifisering. Én metode er å telle leksemer, men det krever et kriterium for å bestemme hvilke ord som er med i bokmål og i SØ. En annen metode er å telle forekomster i løpende tekst, men det krever en enighet om hva slags og hvilke tekster som skal ligge til grunn. Med begge disse metodene risikerer man dessuten at resultatene kan hevdes ikke å være representative, enten ved at en stor mengde infrekvente ord påvirker resultatet i ordboksmetoden, eller ved at noen få frekvente ord med særegen stavemåte påvirker resultatet i korpusmetoden. Et kompromiss ville være å basere undersøkelsen på et visst antall av de mest frekvente ord, men også her vil det være problemer med å bestemme utvalgsriterier. Et problem som er felles for begge metodene, er fastsettingen av hva som faktisk er gjeldende uttalenorm for hvert ord. Som vi skal se, er dette i mange tilfeller et reelt problem. Som vi også skal se i kapittel 7, kan en kvantitativ metode for å bestemme henholdsvis primærfonem og primærgrafem resultere i resultater som strider mot språkbrukernes intuisjon.

En annen mulig definisjon av fonologisk primæruttrykk er det fonologiske uttrykket som vil brukes ved en opplesning der ingen deler av konteksten motvirker realiseringen av dette uttrykket. Men i og med at vi aldri leser bokstaver uten at de står i en kontekst, med unntak av der vi bruker nettopp

bokstavnavnene, er denne definisjonen svært vanskelig å bruke i praksis. Hvis vi ser på et eksempel som <u>, ser vi problemet:

- <lun>=/lu:n/
- <kum>=/kom/

Her har vi at <u> representerer to forskjellige fonemer /u/ og /o/ i to forskjellige kontekster. I dette tilfellet er det den etterfølgende konsonant som i stor grad bestemmer uttalen av <u> i hvert av disse ordene, og om vi ville undersøke dette mer formelt, kunne vi finne flere eksempler på disse omgivelsene for å påvise hvilke deler som har innvirkning på uttalen:

1. <lumpen>=²lompen/, <lunge>=²loŋe/, <lur>=/lu:r/, <lugge>=²luge/
2. <kur>=/ku:r/, <kulse>=²kulse/, <kumle>=²komle/, <kum>=/kom/

Vi ser at det virker lite sannsynlig at det er bokstaven eller fonemet som kommer foran <u>, som bestemmer uttalen av bokstaven i kontekst. Om vi så ser på bokstaven som kommer etter <u>, finner vi:

1. <sunn>=/sun/, <pund>=/pun/, <dun>=/du:n/, <grunn>=/grun/
2. <kum>=/kom/, <nummer>=¹nomer/, <kumpe>=²kompe/,
<skumre>=²skomre/, <dum>=/dom/, men
<sum>=/sum/, <nummen>=²numen/
3. <dunk>=/doŋk/, <lunken>=²loŋken/, <lunge>=²loŋe/
4. <putt>=/put/, <lutt>=/lut/, <kutt>=/kut/, <mutt>=/mut/

Eksemplene viser at det bare er tendenser til at etterfølgende konsonant bestemmer hvilket fonem <u> representerer, men at det i de fleste tilfeller er en ganske sterk tendens. Å definere fonologisk primæruttrykk ut fra uttalen i utlyd vil gi de ønskede resultatene for de fleste vokalgrafemene i SØ-bokmål-relasjonen, men kanskje ikke for <æ>:

- <sa>=/sa:/, <se>=/se:/, <si>=/si:/,
- <bo>=/bo:/, <bu>=/bu:/, <by>=/by:/

<træ>=/tre:/, <trø>=/trø:/, <trå>=/trå:/

For å stemme med tradisjonen må definisjonen av fonologisk primæruttrykk også baseres på opplisting. For vokaler er primærfonemet lik bokstavnavnet. Siden vi regner kvantitet som et suprasegmentalt trekk, trenger vi ikke å spesifisere lengde som del av vokalenes basisuttale. For konsonanter er basisuttalen lik konsonantdelen av bokstavnavnet, med følgende unntak eller presiseringer:

- <c> = /s/, /k/
- <j> = /j/
- <w> = /v/
- <x> = /ks/
- <z> = /s/

Det er vanskelig å insistere på at <c> har ett primærfonem, ettersom uttalen av <c> er så sterkt knyttet til /k/ i mange grafiske eller fonologiske omgivelser. Vi merker oss også at <x> er spesiell ved at den har et primært fonologisk uttrykk som er en *fonemsekvens* i motsetning til et enkeltfonem.

Vi trenger også begrepet primært grafonomisk uttrykk om det grafem eller den grafemsekvens som er sterkest knyttet til hvert fonem. Her støter vi på enda større praktiske og prinsipielle problemer ved bruk av både kvantitative og andre typer kriterier. Primærgrafem, primært grafonomisk uttrykk eller basisstavemåte må også defineres ved opplisting, og ikke alle fonemer har et entydig primæruttrykk. For eksempel er /j/ problematisk her, som vi skal se i kapittel 7.

4.4.2 Polygrafer

I bokmål som i mange andre skriftsystemer er det slik at antall grafemer er mindre enn antall fonemer som skal representeres. Selv om vi bare regner med konsonanter og monoftonger, utgjør disse segmentale fonemene i SØ 30, mens bokmål altså bare har 29 bokstaver, hvorav 5 har et primært fonologisk uttrykk

som er likt andre grafemers (eller grafemsekvensers) primæruttrykk, og som derfor i liten grad bidrar i det fonemografiske systemet.

En metode som er svært vanlig i mange skriftsystemer for å overkomme dette problemet, er å bruke polygrafer. En polygraf er en grafemsekvens som har et primært fonologisk uttrykk som ikke følger av enkeltgrafemenes primærfonem og linearitetsprinsippet, som oftest et enkeltfonem. Eksempler fra vietnamesisk er <ng>=/ŋ/, <ngh>=/ŋ/, <ph>=/f/. Eksempler fra bokmål er digrafen <ng>=/ŋ/ og trigrafen <skj>=/ʃ/. Også diftongenes primære grafemiske uttrykk kan klassifiseres som digrafer: <øy>=/øj/, <ei>=/æj/ og <au>=/æu/, ettersom heller ikke disse sekvensenes basisuttale kan sies å følge fra primærfonemene og linearitetsprinsippet.

I bokmål er det ingen monoftonger som har polygrafer som sine primære grafemiske uttrykk, men teoretisk er det ingenting i veien for det. Historisk sett har <æ ø å> utviklet seg fra ligaturer av to ulike vokalgrafemer, noe som ikke er så ulikt digrafer. I nederlandsk finnes digrafer for monoftonger, for eksempel <oe>=/o/.¹

Noen analyser regner etablerte digrafer som egne grafemer. Vårt strukturalistiske utgangspunkt gjør imidlertid dette vanskelig ettersom vi kan vise ved kommutasjon med null at en digraf ikke kan være den minste innholdsskillende enhet:

<syng>=/syŋ/ : <syn>=/sy:n/

Likevel er det sterke tradisjoner i enkelte skriftspråksamfunn for å se på polygrafer som egne grafemer; for eksempel er vietnamesiske ordlister sortert etter dette utvidede grafembegrep, slik at for eksempel <tu> kommer foran <thu>.²

¹/o/ brukes her for det nederlandske fonemet med artikulasjonssted som ligger nærmest den norske /o/, slik jeg bruker symbolet i denne oppgaven, altså som IPA [u].

²Se for eksempel Nguyen 1991.

Selv om vi ikke regner polygrafer som selvstendige grafemer, inngår de i grafemiske tegn, og vi kan snakke om en polygrafs primærfonem, for eksempel har <ng> i bokmål primærfonemet /ŋ/. Motsatt sier vi at /ŋ/ har det primære grafemiske uttrykket <ng>.

4.4.3 Grafeminventaret

I tillegg til bokstavene slik de er omtalt over, finnes flere andre grupper av grafemer som brukes i bokmål:

- Numeriske grafemer: <0 1 2 3 4 5 6 7 8 9>
- Diakritika: <' ` ~ ^ ~>
- Morfemografiske grafemer: <' ->
- Syntaktiske grafemer: <. ! ? : ; , - () " />
- Andre grafemer: <& § % \$>

Disse gruppene har noe ulik status. I 4.1.2 viste jeg at tallene er semasiografiske enheter. Definisjonen av grafem i 4.2 gjelder bare for skrifttegn i egentlige skriftsystemer, men den kan enkelt utvides til også å gjelde i semasiografiske systemer. Et semasiografisk grafem er den minste betydningsskillende enhet, og vi ser at siffergrafene over må være grafemer. Dette er en lukket mengde.

Den siste gruppen av enheter har litt mindre klar status. <§> er definitivt en utenomspråklig enhet, ettersom reduplikasjon har fonologisk og semantisk innhold som ikke kan representeres språklig i henhold til linearitetsprinsippet:

<§ 1>=*paragraf én*

<§§ 1-3>=*paragrafene én til tre*

<&> er kanskje en morfemografisk enhet, siden den har et klart semantisk innhold ("og"), men kan ha flere ulike fonologiske innhold som alle er uavhengige av det grafiske uttrykket: /å:/, /et/, ?/ænd/.

<%> kan kanskje sies å tilhøre det matematiske notasjonssystemet, men har likevel litt andre egenskaper, i og med at både fonologisk innhold og semantisk innhold er faste: <%>=/pro¹sent/="per hundre". <%> følger også linearitetsprinsippet i tekst, i motsetning til matematisk notasjon, som vi har sett, og <%> kan derfor sies å ha metaskriftlig status, som en forkortelse for <prosent>. Imidlertid er ikke <%> og <prosent> i fri variasjon, slik <f.eks.> og <for eksempel> er; <%> kan ikke brukes i løpende tekst uten i kombinasjon med tall:

*<Han fikk ti % avslag.>

*<Han fikk %er i avslag.>

*<Du må oppgi svaret i %.>

*<Partiet har en framgang på 2 %poeng.>

<\$> har et fast fonologisk innhold ^{/1}dålar/ (i motsetning til i engelsk) og et fast semantisk innhold "entall og flertall av dollar", men det følger ikke linearitetsprinsippet:

<\$1>=/¹e:n ¹dålar/

og må derfor sies å være semasiografisk.

Denne gruppen er en åpen mengde, noe som var tydelig da EU innførte euro-valuta, og det nye grafemet <€> uten problemer ble en del av grafemmengden i bokmål.

At de små grafene som vi kaller diakritika, har innholdsskillende funksjon, ser vi av minimale par som:

<alle>=/²ale/="enhver" : <allé>=/a¹le:/="veg med trær på begge sider"

<for>=/får/=preposisjon : <fôr>=/fo:r/="mat til dyr".

I begge tilfeller må de diakritiske tegnene kontrasteres med null, ettersom disse grafene har svært begrenset distribusjon. Det er svært vanskelig å finne minimale par der de diakritiske grafene kommuterer seg imellom. Kanskje er de alle bare allografer av ett grafem {aksent} i bokmål; distinksjonen har i alle fall svært liten

funksjonsbelastning. Vi ser blant annet av alle kafeer som heter <cafè> og André'er som i avisene omtales som <André>, at grafenes grafemiske status i det minste er tvilsom. De kan kanskje best defineres som kombinatoriske allografer av ett diakritisk grafem, der grafemiske omgivelser bestemmer hvilken allograf som passer. Imidlertid finnes det ett eksempel på minimale par med utvilsomme bokmålsord, og dette gjør det vanskelig å se bort fra i hvert fall denne distinksjonens innholdsskillende funksjon:

<fôr>=/fo:r/="dyremat" : <för>=/få:r/="plogrenne" (sideform til <får>)

Dessuten signaliserer de forskjellige diakritiske grafene ulik endring i fonologisk uttrykk.

Eksempler fra Vinje 2004 er <allé>, <òg>, <fôr>, <Citroën>, <señor>, <provençalsk>, og dette er alle de diakritika han omtaler. Språkrådet oppgir imidlertid på nettsidene sine¹ en liste over utenlandske geografiske navn der mange skal ha andre typer diakritika i bokmålstekster, til og med diakritika som ikke finnes på vanlige norske PC-tastatur. Jeg har ikke undersøkt om det blant disse geografiske navnene kan finnes minimale par ved substitusjon av diakritika, men på grunnlag av dette må vi også si at mengden av diakritika er åpen, enten de defineres som grafemer eller allografer.

Det kan reises spørsmål om hvorvidt ikke <°> og </> er diakritika i bokmål, og at de har vokalkvalitetsendrende funksjon på henholdsvis <o> og <a>. Historisk sett har disse utvilsomt utviklet seg fra diakritika, og tilsvarende vokalkvalitetsendrende grafer regnes ofte som diakritika i andre skript, for eksempel <¨> i tysk. Slik oppfattes de ganske sikkert også av utlendinger uten inngående kjennskap til bokmål. Imidlertid er det to forhold som taler imot en slik analyse. For det første opptrer disse grafene bare i kombinasjon med ett grafem hver; vi har ikke noe <û> i bokmål. For det andre kan vi ikke se bort fra tradisjonen og den allmenne oppfatningen blant skriftspråksbrukere om at <å> og <ø> er separate grafemer som blant annet har sin egen plass i alfabetet og i ordlister.

¹<http://www.sprakrad.no>

I gruppen av morfemografiske grafemer finnes apostrof og bindestrek.

Apostrof brukes hovedsakelig som en markør for eieform etter <s x z> der denne ikke skiller seg fonologisk fra grunnformen:

<Hans'>=/hans/="Hans sin"

Men den brukes også som en markør for eieform etter <s x z> selv om eieformen skiller seg fonologisk fra grunnformen:

<Dumas'>=/dy¹mas/="Dumas sin"

I begge tilfeller fungerer apostrofen morfemografisk: <'>=/Ø/, /s/="genitiv¹ av ordet foran". Alternativt kunne apostrofen her også analyseres som en metaskriftlig representasjon av <s> i visse omgivelser:

<Dumas'>=<Dumass>

<Hans'>=<Hanss>

Imidlertid forutsetter denne analysen at det metaskriftlige uttrykket representerer et grafisk uttrykk som ikke er del av bokmål, og dette er i strid med definisjonen av metaskrift i 4.3.4.

Apostrof har også andre bruksområder², men disse skal vi ikke analysere her.

Bindestrek har også mange til dels beslektede bruksmåter som vi ikke kan gå inn på i detalj, men hovedprinsippet er å vise plassering av morfemgrensen i en sammensetning: <trafikk-kork>, <pc-en>, <50-øring>, <svart-hvitt>. I noen tilfeller indikerer sammensetning en annen uttale enn om ordet ikke hadde vært sammensatt:

<sne-is>=/²sne:i:s/="is med sne-karakter" : <sneis>=/snæjs/="pinne"

¹Det er nå relativt stor enighet om at SØ ikke har genitiv, og at en analyse av dette som genitiv må støtte seg på diakrone analyser. Den grammatiske analysen av dette er imidlertid ikke viktig her; apostrofen er en parallell til <s> i <Påls> eller <guttens> uavhengig av hvilken grammatisk status vi gir den.

²Se Vinje 2004.

men det er den morfologiske markeringen som er primær, mens endringen i fonologisk uttrykk er sekundær og en konsekvens av morfemgrensen.³ Vi kan derfor kalle bindestrek for et morfemografisk grafem.

Mellomrom, som kan realiseres også som linje- eller sideskift, kan også klassifiseres som et morfemografisk grafem, ettersom funksjonen er å skille mellom ordformer i en tekst. Det grafiske mellomrommet trenger ikke ha noen fonologisk lisensiering fra den talte ytringen, men i mange tilfeller vil mellomrommet være lisensiert av hovedtrykk i begge de skilte ordformene eller av andre fonologiske forskjeller:

<lamme lår>=/²lame¹lâ:r/="lår med lammelse" :

<lammelår>=/²lamelâ:r/="lår fra lam"

<dersom>=/¹dæfâm/="hvis" : <der som>=/¹dæ:fâm/="hvor"

De syntaktiske grafemer har også viktige semantiske og lesestøttende funksjoner, men disse faller utenfor denne diskusjonen. Se ellers om skilletegn i 4.4.3.

4.5 Fonemografi

Det bærende prinsipp i fonemografiske skript er en en-til-en-relasjon mellom fonem og grafem. At den er én-til-én innebærer at den er entydig begge veger: hvert fonem lisensierer ett og bare ett grafem, og ett grafem representerer ett og bare ett fonem. Som vi har sett, gjelder dette ikke gjennomført for bokmål i dets relasjon til SØ, og det er tvilsomt om det finnes noe egentlig skriftsystem der dette gjelder unntaksløst. Prinsippet kan brytes på flere måter:

- ett fonem kan lisensiere ulike grafem, f.eks. /ån/=<ånd>, /ån/=<onn>,
- ett grafem kan representere ulike fonem, f.eks. <ond>=/on/, <onn>=/ån/,
- ett fonem kan lisensiere en grafemsekvens, f.eks. /çe:/=<kje>,
- ett grafem kan representere en fonemsekvens, f.eks. <taxi>=/¹tæksi/,

³Man trenger også en del fantasi for å finne gode minimale par her!

- et fonem eller en distinksjon lisensierer ikke noe grafem, f.eks. / ¹bedre/, / ²bedre/= <bedre>,
- et grafem representerer ikke noe fonem, f.eks. <hva>=/va/ i bokmål eller <h>=/Ø/ i spansk.

Det er også mange årsaker til disse avvikene fra prinsippet, og det er disse årsakene som er tema for dette kapitlet.

4.5.1 Linearitet

Tale kan sees på som et fenomen som er lineært i tid. Enhver talt ytring har en utstrekning i tid, og om et kommunikasjonsteoretisk informasjonsbegrep legges til grunn, kan man si at all informasjon i en talt ytring ligger i én dimensjon langs en tidsakse. Uansett om man tolker informasjonsbegrepet smalt som semantisk innhold eller bredt som semantisk innhold pluss pragmatikk, språkfunksjoner og akustiske signaler med ekstralingvistisk innhold, som tonegang, volum, stemmekvalitet og eventuelt andre variabler, bæres informasjonen av endringer i det akustiske signalet langs tidsaksen. Ikke-akustiske signaler som mimikk og fakter kan neppe sies å ligge innenfor det én-dimensjonale her, men slike fenomener faller også utenfor denne oppgavens diskusjon om skrift.

De fleste betydningsskillende enheter i talespråk er slike som gjerne omtales som segmentale enheter, og denne omtalen harmonerer med oppfattelsen av tale som lineær. Imidlertid er det to ulike typer problemer forbundet med dette synet:

1. Taleorganene er i kontinuerlig bevegelse, slik at det verken fysiologisk eller akustisk sett finnes diskrete, segmentale enheter i talen. Det akustiske signalet består tvert i mot av variabler som er i kontinuerlig endring. Dette medfører at trekk fra ett fonem gjerne smitter over på det etterfølgende fonemet og/eller på det foregående fonemet. De såkalte segmentale enhetene er dermed ikke fysisk diskrete enheter.
2. Talespråk består i tillegg til de segmentale enhetene også av enkelte prosodemer. Disse enhetene blir gjerne kalt *suprasegmentale*, fordi deres

akustiske egenskaper kommer i tillegg til og samtidig med de segmentale enhetene, og de opptrer gjerne over en sekvens av mer enn én segmental enhet. De kan dermed sees på som ekstra dimensjoner i forhold til det én-dimensjonale segmentale nivået.

Det er viktig å merke seg skillet mellom fysikk og funksjon her. Begrepene segmentale og suprasegmentale enheter anvendt på talte ytringer er funksjonelle *tolkninger* som søker å strukturere de fysiske signalene. Tolkningene kan gjerne innebære at man ser på talen som todimensjonal, der det segmentale og det suprasegmentale utgjør hver sin dimensjon. Signalene kan også tolkes som flerdimensjonale, der det segmentale og hvert av de suprasegmentale trekkene utgjør én dimensjon, noe som ville innebære at ytringer i SØ ville være firedimensjonale. Uansett på hvilken måte man velger å strukturere talen, ligger det lineære akustiske signalet til grunn.¹

Når det gjelder skrift, kan den også sees som lineær. Dette kan begrunnes i tre ulike perspektiver, som legger vekt på ulike aspekter ved skriften:

1. Hvis skrift er definert som en representasjon av tale, og tale er lineær, må skrift også være lineær.
2. Når skrift persiperes, blir den lest i en viss rekkefølge, altså lineært i tid.
3. Det grafiske uttrykket viser tydelig skriftens lineære natur; skrift er visuelt og fysisk lineær.

Alle perspektivene over har viktige elementer i seg, men de har også viktige problemer forbundet med seg:

1. Dette er et svært enkelt perspektiv, og det er enkelt å tilbakevise med et moteksempel. Også grafisk utskrift av lydbølger *representerer* tale, men en slik utskrift er todimensjonal. Om dette perspektivet på skrift skal holde som en definisjon og et bevis på at skrift er lineær, må begrepet

¹ Også en grafisk fremstilling av lydbølger er todimensjonal. Da er det signalets volum på mikronivå som en funksjon av tid man representerer. Dette er mindre relevant for denne oppgaven.

"representere" defineres mer presist. Likevel er det klart at dette perspektivet inneholder et viktig poeng; om skrift skal representere tale på en funksjonell måte, må den kunne *fungere* lineært.

2. Når en leser skal lese en tekst som er trykket eller skrevet på papir, er det normalt ingenting som hindrer ham i å velge en annen rekkefølge enn den rekkefølgen som skribenten har intendert, eller den rekkefølgen som er indikert av trykksakens design, og teksten kan således oppfattes som fler-dimensjonal. Mange elektroniske tekster har ikke de samme egenskapene som trykte eller skrevne tekster når det gjelder dette perspektivet; det er svært vanlig at elektroniske tekster er designet nettopp for å underbygge leserens valgfrihet med hensyn til leserekkefølge.¹ Imidlertid gjelder dette også mange trykte tekster, der strukturbyggende fasiliteter som for eksempel innholdsfortegnelse, register og fotnoter i vitenskapelige tekster, og bildetekster og "faktabokser" i tabloidaviser, oppfordrer til alternative lese måter. Det som imidlertid gjenstår som en viktig og riktig essens i dette perspektivet, er at den *leste* teksten – altså det som leseren faktisk leser – nødvendigvis må være lineært. Lesingen konstruerer en lineær ytring.
3. Om man ser på et ark med en trykt tekst, ser man øyeblikkelig at skriften ikke er fysisk én-dimensjonal. Den er faktisk to-dimensjonal på to nivåer. For det første har hver enkelt bokstav en utstrekning i to dimensjoner,² og for det andre er bokstaver i en tekst av en viss lengde fordelt todimensjonalt utover arket. Faktisk vil bokstaver i en tekst av en viss lengde være fordelt tredimensjonalt utover et boklig rom, i tillegg til eventuell dimensjons-økende funksjonalitet som beskrevet i over.

¹ Enkelte elektroniske tekster er også designet med det motsatte for øye, nemlig å pålegge leseren én leserekkefølge. Tekniske stengsler sørger for at teksten kan vises kun i én sekvens og slik sett er strengere lineær enn ordinære trykte tekster.

² Et unntak fra dette er morsesignaler som er nedskrevne. Slik skrift er fysisk lineær, og dette er selvfølgelig fordi den er en svært naturtro transskripsjon av de en-dimensjonale strømsignalene som er morsens primære fysiske uttrykk. Som det går fram av 4.3.4, ser vi på morse som metaskrift.

Imidlertid er ikke fokus i dette perspektivet på skriftens fysiske egenskaper, men dens logiske. En skrevet ytring er logisk sett lineær.

Når det gjelder segmentalitet, er kanskje situasjonen enklere i skrift enn i tale. I hvert fall i trykt skrift er grafemene som regel tydelig atskilt av hvitt rom og dermed segmentale, både logisk og fysisk. Det finnes riktignok visse mulige unntak. For det første har vi *ligaturer*, som gjerne tolkes som to bokstaver som er samskrevet, for eksempel som i fransk <manœuvre>. Også de norske bokstavene <æ ø å> er historisk sett ligaturer, men etter en funksjonell analyse av det norske skriftsystemet, er de definitivt nå separate grafemer. For det andre har vi det typografiske knepet *kerning*, der visse bokstavsekvenser blir trykket tettere enn normal satsbredde skulle tilsi for å øke lesbarheten, og som i visse tilfeller medfører at bokstavene henger sammen, for eksempel som i <fi> i stedet for <fī>. Ligaturer brukes normalt ikke i moderne bokmål, og vi skal ikke se nærmere på det. Kerning tilsvarende både fysisk og logisk håndskrift, som vi behandler under.¹

I håndskrift vil det ofte være et segmenteringsproblem, tilsvarende som i tale. Bokstavene henger sammen fysisk uten noen klar grense som viser hvor én bokstav ender og den neste begynner. I enkelte tilfeller vil den etterfølgende bokstav begynne lenger til venstre (i et venstre-til-høyre-skript som bokmål) enn den foregående bokstav ender, og kanskje til og med begynne lenger til venstre enn den foregående bokstav *begynner*, som for eksempel i <hig>, men funksjonelt er det normalt ikke vanskelig å avgjøre hvilken rekkefølge enhetene opptrer i og skal tolkes i.

Imidlertid er det en annen type grafemer som er vanskeligere forenbare med skrift som et lineært fenomen, nemlig diakritiske tegn som <' ^> i for eksempel <allé> og <fôr>. Disse grafene opptrer ikke lineært på linje med de segmentale grafemene, men todimensjonalt. Denne todimensjonaliteten er ikke direkte

¹ Typografihistorisk har ligaturer og kerning sannsynligvis felles opprinnelse.

relatert til prosodemenes flerdimensjonalitet i talen, selv om i hvert fall én, <'>, har en relasjon til trykkplassering i et ord som <allé>=/a¹le:/.

Om vi ikke definerer versaler og minuskler som ulike grafemer, må de skilles ved et betydningsskillende trekk {versal}. Dette trekket kan tenkes å stå til et segmentalt grafem og modifisere uttrykket til dette grafemet:

$$<\{\text{versal}\}\text{hans}> \rightarrow <\text{Hans}>.^1$$

Dette trekket opptrer utvilsomt utenfor det endimensjonale, selv om det resulterende uttrykket er endimensjonalt. Dette minner også om Chomsky og Halles² teori om dypfonologi og overflatefonologi, der enheter i dypstrukturen ikke i seg selv kommer til uttrykk på overflaten, men påvirker overflatestrukturen på ulike vis. En slik tolkning kunne kanskje kalles dypgrafemisk.

4.5.2 Retning

Tale og skrift er to ulike tegnsystemer. Det eksisterer flere typer av korrespondanser mellom disse systemene, og for noen av disse typene er det relevant om vi ser på den som en relasjon fra tale til skrift eller en relasjon fra skrift til tale. Relasjonen fra skrift til tale sier vi er orientert i *leseretningen* eller *persepsjonsretningen*, mens relasjonen fra tale til skrift sier vi er orientert i *skriveretningen* eller *produksjonsretningen*.

Fonem-grafem-korrespondanser tilhører skriveretningens problemstillinger. For eksempel kan fonemet /ʃ/ lisensiere <s>, <sj>, <skj>, <sk>, <g> og <rs> i hjemlige ord; i enkeltstående tilfeller også <tsj> og <sgj>. I mer fremmedartede importord er i tillegg <ch> og <j> aktuelle.

Grafem-fonem-korrespondanser tilhører leseretningens problemstillinger. For eksempel kan grafemet <o> representere <o> og <å> i hjemlige ord.

¹Eller til bokstavparet.

²Chomsky og Halle 1968.

4.5.3 Nivå og skopus

Det er flere lingvistiske nivåer som er relevante for en omtale og en diskusjon av bokmåls forhold til SØ. De viktigste er ordet og morfemet på den ene siden, og stavelsen og fonemet på den andre.

Både fonotaktiske, grafotaktiske og grafonomiske regler opererer innenfor et virkeområde; de har et definert skopus. Viktige avgrensningsmekanismer her er ordgrense, morfemjunktur og stavelsesgrense. Om en regel gjelder "frontalt" kan dette med andre ord bety ordfrontalt, morfemfrontalt eller stavelsesfrontalt, og dette må enten være uttrykt eksplisitt eller gå fram av sammenhengen. Tilsvarende gjelder for "finalt".

Der en regel har skopus innenfor et morfem, vil det ofte være relevant om det er et rotmorfem (eller en stamme), et avledningsmorfem eller et bøyningmorfem. Om regler som har morfemjunktur som én av sine variabler, vil det være relevant om det er en sammensetningsjunktur, en avledningsjunktur eller en bøyningjunktur. Også hvorvidt den morfologiske mekanismen er produktiv, vil ofte være en relevant variabel.

4.5.4 Markørgrafemer

Venezky 1970 kategoriserer grafemer i engelsk som enten fonografiske eller *markører*. Det er som kjent store forskjeller mellom norsk og engelsk grafonomi, men denne angrepsvinkelen kan til dels også brukes for bokmål, med visse tilpasninger. De fonografiske grafemene er grafemer som representerer et fonologisk uttrykk i det aktuelle grafonomiske tegnet, mens markørenes primære funksjon er å modifisere uttalen, å opprettholde et morfologisk mønster eller å markere leksikalsk tilhørighet.

Eksempler:

1. I <mulig>=²mu:li/ er <muli> fonografiske, mens <g> er en grammatisk markør for adjektiv eller adverb, altså en kategorimarkør.

2. I <eplet>=/¹eple/ er <eple> fonografiske, mens <t> er en morfologisk markør for bøyningen {bestemt form entall nøytrum}. Men i <eplets>=/¹eplets/ er alle grafemene fonografiske, selv om de samtidig representerer morfologisk innhold. Venezky anser skillet mellom fonografi og markører for å være en dikotomi, slik at grafemer med fonologisk innhold ikke samtidig kan sies å være markører, selv om de også skulle ha andre funksjoner enn den rent fonemografiske.
3. I <tolv> er <tol> fonografiske, mens <v> en leksikalsk markør som skiller <tolv>=/tål/="12" fra <toll>=/tål/="avgiftstype".
4. I samme ord er <v> også fonologisk markør, i og med at <lv> som konsonantsekvens markerer at vokalen er kort.¹ Av kvantitetsmarkeringen følger også en kvalitetsmarkering, nemlig at <o>=/å/, som i <toll>=/tål/, <vold>=/vål/, i motsetning til <bol>=/bo:l/, <pol>=/po:l/. Om det er kvantitetsmarkeringen eller kvalitetsmarkeringen som er primær her, er kanskje usikkert, men om <v> ikke hadde vært til stede i ordformen, måtte ordet hatt en annen konsonant i samme posisjon for å oppfylle kvantitetsregelen

{kort vokal}|_KK ellers {lang vokal}

5. I <hvor>=/vor/ er <vor> fonografiske, mens <h> er en semantisk markør for spørreord.²

Alle tolkningene over er problematiske. Ingen av dem er absolutte, så de representerer uansett bare tendenser og ikke regler i bokmål.

1. De fleste spørreord med /v/ i fremlyd har #<hv>. Om vi bruker en smal definisjon av spørreord, gjelder det alle: <hva>, <hvem>, <hvor>, <hvorfor>, <hvordan>, <hvilken>, mens <hvis>, <viss> og <verken>³

¹<v> i <tolv> markerer også en morfologisk sammenheng med <f> i <tylft>. Denne forbindelsen er imidlertid hovedsakelig nå av historisk karakter.

²Spørreord kan også sees som en kategori, og da er <h> en grammatisk markør, som i (1).

³ Begge disse ordene hadde tidligere eneformer med <hv>. <hverken> er for tiden til vurdering.

faller utenom om vi regner dem som spørreord. Imidlertid finnes det mange ord som definitivt ikke er spørreord, med #<hv>, for eksempel <hval>, <hvit>, <hver>.

2. Regelen over er for vid og slipper gjennom substantiver som <sig> og <krig>. En regel basert på antall stavelser, der antall stavelser må være større enn én for at regelen skal være anvendbar, ville være betydelig mer effektiv, men ville kunne slippe gjennom mulige sammensetninger med substantiver på <ig># som sisteledd, for eksempel <tilsig> eller <borgerkrig>. Vi ser at regelen må baseres på et kriterium som er knyttet til avledningsmorfemene -/i/ og -/li/, for å kunne fungere uten unntak i persepsjonsretningen. Det er imidlertid mest trolig at den psykologiske realiseringen av en slik persepsjonsregel heller er basert på en hovedregel formulert som i

$$\langle g \rangle \# = \text{trykklett } i/\emptyset/$$

med visse unntak, gitt den relative frekvens som disse forekomstene har.

3. Det er problematisk at <t> og <g> i eksemplene over veksler mellom å være markører og fonografiske enheter ettersom de er fonologisk realisert eller ikke. Vi har også at <et>|_# = /et/ i andre kategorier, for eksempel <funnet>=/²funet/.
4. <lv># = /lv/ i mange ord, f.eks. <gulv>=/gulv/, mens l<v> i andre representerer en morfofonemisk veksling, f.eks. i <halv>=/hal/ ~ <halvere>=/hal¹ve:re/.

Følgende presiseringer og spesifiseringer kan være nyttige.

Markører opererer innenfor en G/F-kontekst, altså en grafonomisk og fonologisk kontekst. De er således ikke autonome enheter i en perspepsjonssituasjon, i og med at det ikke er noen sekvens av operasjoner leseren kan gjøre på det grafiske uttrykket for å bestemme hva som er fonografiske enheter og hva som er markører. De fungerer i et samspill i de tovegs relasjonene som finnes mellom

grafisk og fonologisk uttrykk. I det grafiske uttrykket <vannet> kan <t> være en fonografisk enhet i /²vanet/, men en markør for kategori og morfologisk informasjon i /¹vane/, og leseren må bruke syntaktisk informasjon, og kanskje informasjon på et høyere tekstlig nivå, for å avklare den fonologiske konteksten som bestemmer om <t> er en markør eller en fonografisk enhet.

En *svak markør* for en egenskap er en markør som i identiske grafonomiske eller fonologiske omgivelser også kan markere andre egenskaper. <^> i <fôr> er en svak leksikalsk markør for "dyremat" fordi den også kan markere for "tøylag" i den samme G/F-konteksten.

En *partiell markør* er en markør som ikke markerer for alle tilfellene av den aktuelle forekomsten. For eksempel er <v>=/Ø/ i <gav>=/ga:/ en rotmorfemmarkør for roten i ord som <gave>=/²ga:ve/, <giver>=/²ji:ver/, <giv>=/gi:v/, men den er en partiell markør for dette rotmorfemet ettersom den ikke er til stede i <gi>=/ji:/ eller <gitt>=/jit/.

En *multifunksjonell markør* er en markør som markerer for flere ulike egenskaper i samme ord. For eksempel er <d> i <godt> en leksikalsk markør som markerer identitet i rotmorfemet mellom /go:/ og /gå:/, men den er også en fonologisk markør som – fordi den sammen med <t> utgjør en grafemsekvens – signaliserer kort vokal.

Venezky opererer altså med en dikotomi mellom fonografiske enheter og markører basert på hva som er grafemets eller grafemsekvensen *primære funksjon*. En slik dikotomi er kanskje litt rigid, og det er antagelig mer formålstjenlig i bokmål å snakke om grafemenes fonografiske funksjoner og markørfunksjoner. For eksempel markerer <eg>=/æj/ i <jeg> svakt og partielt for kategorien personlig pronomen, men både grafemsekvensen og <g> sett i isolasjon er nødvendige for å representere det fonologiske uttrykket. Verken <j> eller <je> kan representere "jeg". Tilsvarende rolle har <g> i <gi>, som markerer samme rotmorfem som i <gav>, <gave>, men som også har en fonemografisk rolle. Det blir mer presist om vi sier at grafemer som i en G/F-kontekst avviker

fra linearitetsprinsippet, har en mer eller mindre uttalt markørfunksjon i tillegg til den fonografiske funksjonen.

At markørene opererer i en fonologisk kontekst, kan ha to ulike fokus. /¹seksi/= <sexy> kan ikke få <g> som i * <sexyg> fordi den grafonomiske konteksten er feil, altså på grunn av <y>. Men sett i en fonologisk kontekst kan vi heller ikke ha /¹seksi/= * <sexig> på grunn av feil tonem. <ig># i denne morfologiske konteksten korresponderer nemlig med tonem 2, som i <herlig>= /²hæ:li/, <farlig>= /²fa:li/. Dette kan analyseres på to ulike måter:

(a) <g> er også en fonologisk markør for tonem 2.

(b) Tonem 2 er del av nødvendig fonologisk kontekst for å bruke <g> som morfologisk markør av adjektiv.

Fokus i (a) er på persepsjonsretningen, mens fokus i (b) er på produksjon. Det er nok en sterk tendens til at leseretning er dominerende her; <sexig> → /²/.

Stavemåten vil generere lesinger med tonem 2.

Tilsvarende kan vi prøve å analysere <t> i <skrivet>= /¹skri:ve/ som en markør av tonem 1, til forskjell fra <skrive>= /²skri:ve/. De to utgjør et minimalt par med tonem i kommutasjon. Det samme gjelder <vanne>= /²vane/ : <vannet>= /¹vane/, men her blir situasjonen komplisert ved homografen <vannet>,v= /²vanet/, som ikke er et minimalt par med <vannet>,s. I <vannet>,v er <t> derimot *ikke* en markør, ettersom <t>= /t/ i denne ordformen, og da stemmer det for såvidt godt at denne ordformen også har tonem 2. Imidlertid er dette bare en følge av tonemets rolle i bøyings- og avledningsmorfologien i SØ, og det er ikke en regelmessig funksjon av <t> som markør, som vi kan se av eksempler som <eplet>= /²eple/. Det er ingen regelmessige markører av tonem i bokmål. Se 4.5.5 og 5.6.2.

4.5.5 Dekningsgrad

Sproat 2000 bruker et eksempel fra russisk for å belyse begrepet *dekningsgrad*¹. Et skripts fonologiske dekningsgrad dreier seg om hvilke fonologiske enheter skriptet representerer, og Sproat viser med homografene

<города>=/¹gorədΛ/="en bys" : <города>=/gərΛ¹da/="byer"

at russisk skript ikke dekker trykkplassering. Tilsvarende har vi at bokmål normalt ikke dekker prosodemene trykkplassering og tonem, mens skriptet har partiell dekningsgrad med hensyn til kvantitet.

1. Trykkplassering er normalt ikke markert i grafiske uttrykk, for eksempel <lever>=/¹le:ver/ : <lever>=/le¹ve:r/. Imidlertid finnes det ordformer der trykkplassering på siste stavelse i visse tilfeller lisensierer grafemet <'>, som posisjoneres vertikalt over bokstaven som representerer stavelseskjernen, for eksempel <alle>=/²ale/ : <allé>=/a¹le:/. Dette er imidlertid ikke systematisk og må heller klassifiseres som leksikalsk markering av visse leksemer.² Det er teoretisk mulig å bruke dette grafemet til å markere trykk i ordformer mer systematisk, men dette er ikke en del av bokmål. I spansk, derimot, markeres trykk entydig av <'>. I visse ordformer kan kanskje konsonantreduplikasjon sies sekundært å markere trykkplassering i G/F-konteksten siste stavelse, som i <ballet>= [¹--] : <ballett>=[¹-], men dette er bare en partiell markør, jf. <tunnel>= [¹-].
2. Tonem kan aldri utledes direkte av grafiske uttrykk, for eksempel <banner>=/²baner/ : <banner>=/¹baner/. Indirekte kan tonem utledes med stor statistisk sikkerhet i visse omgivelser, særlig ved <e>#, for eksempel <vanne>=/²vane/ : <vannet>=/¹vane/, men dette er både en partiell og en

¹ Sproat skriver på engelsk og bruker termen *coverage*.

² Egentlig er dette en leksikalsk markering av ordformen, ettersom rotmorfemet bare i visse omgivelser har <'>, og ikke i <alleen>=/a¹le:en/ og <alleer>=/a¹le:er/.

svak markør, som kan sees i for eksempel <vannet>=²vanet/,
<eplet>=²eple/, <færre>=¹fære/ og <verre>=¹være/.

3. Kvantitet er prinsipielt dekket i bokmål ved at i den fonologiske konteksten trykksterk stavelse markeres kort vokal ved etterfølgende konsonantgrafemsekvens, f.eks. <vilt>=/vilt/. Konsonantsekvens kan imidlertid bare være en *svak* markør, ettersom vi har eksempler som

<avgud>=²a:vgu:d/,
<åpning>=²å:pniŋ/,
<bilt>,v=/bi:lt/, <fult>=/fu:lt/
<sart>=/sa:t/,
<påske>=²på:ske/.

De fire første eksemplene forklares ved at regelen har skopus som er avgrenset av morfemgrense: *av+gud*, *åp+ning*, *bil+t*, *ful+t*. *Sart* og *påske* kan imidlertid ikke forklares med regelskopus.¹

En sterkere markør for kort vokal vil være *konsonantreduplikasjon*, altså en sekvens av to like konsonantgrafer, som i <ball>=/bal/, <ballett>=[⁻¹let], <fullt>=/fult/. Tilsynelatende avvik, som <avvik>=²a:vik/, forklares igjen ved at regelen har skopus som er avgrenset av morfemgrense. Imidlertid viser eksemplene under at konsonantreduplikasjon også innenfor dette skopus er en partiell markør:

<til>=/til/,
<øl>=/øl/²,
<lever>=¹lever/,
<kom>=/kåm/.

¹*Sart* kan forklares ved at digrafen <rt> fungerer som én funksjonell enhet. Se kapittel 4.5.8 under.

²Krogsrud og Seip 1942 sier at <øl>, <smør> og enkelte andre ord har enkeltkonsonant på grunn av "vaklende uttale". Jeg er ganske sikker på at uttale med kort vokal i disse ordene er ganske enerådende i dagens SØ, og NUOV er enig med meg. Funksjonsord som <til>, <man> sier de har enkeltkonsonant fordi de sjelden er aksentuerte og dermed sjelden har lang konsonant.

<legning>=^{/2}legnin/,

<mandag>=^{/1}manda/

Det ser ut til at det er vanskelig å finne en markør for kvantitet som er både sterk og total, og dette er jo også et område av bokmål som volder både barn og voksne mye problemer.¹

Ett annet problem som må tas hensyn til ved vurdering av dekningsgrad av vokalkvantitet, er det faktum at trekket er distinktivt kun i trykksterke stavelser, og trykkplassering er ikke dekket i bokmål, som vi så over. Det vil si at i ordformer med flere enn én stavelse, er ikke kvantitet dekket i en grafonomisk kontekst; man må kjenne den fonologiske konteksten for at noen av de potensielle reglene over skal virke.

Sproat bruker bare begrepet dekningsgrad i forbindelse med fonologi, men i et morfemografisk skript eller et blandingsskript kan man gjerne også snakke om morfologisk dekningsgrad. Om vi regner bokmål som et blandingsskript som for eksempel representerer {bestemt form entall nøytrum} ved <t>, der bokstaven ikke er lisensiert av noen fonologisk enhet, altså <t>=/Ø/, har vi at skriptet *dekker* den morfologiske enheten {bestemt form entall nøytrum} også når dette opptrer som et nullmorfem, mens skriptet ikke dekker {ubestemt form flertall nøytrum}=/Ø/ der dette morfemet er et nullmorfem.

<eple>=^{/2}eple/ + <t>={+bestemt}

<hus>=/hu:s/ + <Ø>={+flertall}

Skriptet har altså en partiell dekningsgrad med hensyn til nullmorfemer i nøytrumssubstantiv; det dekker bestemthet (totalt) men ikke flertall.

Til slutt må vi si at bokmål har en mekanisme for dekning av morfemgrense, nemlig <->, som *kan* benyttes som en persepsjonsfremmende markør ved behov. Skriptet har en *potensiell* dekning av morfemgrense. Det vil si at <-> er en

¹Se Wiggen 1992 om barns rettskrivningsavvik. Et raskt søk i Usenet-korpuset avslører at omkring 10% av forekomstene av *spilte* og *spilt* er feilstavet med dobbel <l> (817 mot 7567), mens noe færre (ca. 6% eller 904 mot 15 738) har feilstavet *fullt* med enkel <l>. Dobbelt <n> ser ut til å volde mindre problemer.

partiell markør av "morfemgrense ved behov", men at bruk av <-> som en svak markør, altså av "morfemgrense som ikke er vanskelig å persipere", ikke er del av skriptet bokmål.

4.5.6 Dybde

Begrepet dybde er vanlig i generative fonologiske beskrivelser,¹ men brukes for eksempel av Sproat 2000 om i hvilken grad et skript representerer fonologisk overflatestruktur eller "dypstruktur", altså i hvilken grad og på hvilken måte skriptet representerer regelmessige morfofonemiske vekslinger. Eksemplet fra russisk over som ble brukt til å illustrere dekningsgrad, kan brukes også til å illustrere dybde.

Realiseringen av vokalene i <города> avhenger av hvilken stavelse trykket er på, og vekslingen er automatisk og en funksjon av trykkplasseringen. Vi kan si at en dyp representasjon av ordets fonologi er {goroda}, og at hver av disse fonologisk dype enhetene realiseres som en fon som er medlem av et overflatefonem² etter hvilken plassering de har i forhold til stavelsen med hovedtrykk. <города> er dermed en regelmessig representasjon av den dype formen {goroda} der hvert dypfonem er representert ved sitt primærgrafem:

{g}=<r>

{o}=<o> → /o/|¹₋, /Λ/|¹₋, /ə/|ellers

{r}=<p>

{d}=<д>

{a}=<a> → /a/|¹₋, /Λ/|ellers

¹For eksempel Katamba 1989, Durand 1990, Chomsky og Halle 1968.

²Dette er ikke en presis gjengivelse av noen vanlig fonologisk teori innenfor den generative skolen.

Generative fonologer opererer normalt med et noe annet fonembegrep enn strukturalistene, jf.

Kristoffersen 2000, der de dype enhetene er mer abstrakte enheter enn strukturalistiske fonemer, mens overflateenheter har en posisjon som er nærmere fysiske enheter – fysiologiske, akustiske eller auditive. Også ulike strukturalister opererer imidlertid med ulike abstraksjonsnivå for fonemene, men de er gjerne plassert mellom generativistenes dype enheter og overflateenheter. Man kan si at min gjengivelse over er en slags strukturalistisk tilpasset generativ fonologi, noe som passer formålet i denne oppgaven.

I SØ er det ikke så vanlig med automatiske vekslinger i vokalismen. Avlyds- og omlydsfenomener er ikke så regelmessige i moderne norsk at man kan lage slike regler for dem, men det finnes andre typer tilfeller i bokmål som tydelig viser eksempler på dyp grafonomi.

I bøyningsmønsteret for verbet *gi* ser vi:

<gi>=/ji:/

<gir>=/ji:r/

<gav>=/ga:/¹

<gitt>=/jit/

Primærfonemet for <g> er /g/, men vi ser at <g> i denne stammen representerer både /g/ og /j/. Vi kan her si at formene med /j/ har en annen dypstruktur enn de fonemsekvensene som vi ser over, nemlig {gi:}, {gi:r}, {gitt}, og det er disse dype formene som er representert ved sine primærgrafemer. Uttalen, den fonologiske overflateformen, blir så generert etter en enkel regel:

$$\{g\} \rightarrow /j/ _ V \in \{i, y\}, /g/ _ V \notin \{i, y\}$$

(Venezky kaller de dype enhetene *morfofonemiske enheter*, og viser, som Chomsky og Halle 1968, at det i stor grad er disse enhetene engelsk rettskrivning gjengir. Deres bruk av termen *morfofonemisk* er ikke i samsvar med med den strukturalistiske bruken, slik den blant annet framgår av Haslev 1972.)

I bøyningsmønsteret for verbet *skyte* ser vi den samme mekanismen:

<skyte>=²ʃy:te/

<skyter>=¹ʃy:ter/

<skjøt>=/ʃø:t/

<skutt>=/skut/

¹Vi ser bort fra <v>=/Ø/ i dette eksemplet.

Her ser vi at <sk> representerer både /sk/ og /ʃ/ i bøyningsmønsteret. Vi kan sette opp dypformene {sky:te}, {sky:ter}, {skø:t}, {skut} og generere de fonologiske overflateformene med en like enkel regel:

$$\{sk\} \rightarrow /ʃ/_V \in \{i, y, \emptyset\}, /sk/_V \notin \{i, y, \emptyset\}$$

Dette genererer de riktige fonologiske overflateformene, men det grafiske uttrykket er ikke en direkte gjengivelse av dypfonemenes primærgrafem i dette leksemet, noe som ville gitt <skyte>, <skyter>, *<skøt>, <skutt>. Det er altså ikke noe gjennomført prinsipp i bokmål at det grafonomiske uttrykket representerer de dypfonologiske formene i denne typen vekslinger foran fremre vokaler. For å generere de riktige grafiske formene for dette leksemet, trenger vi en alternativ generativ regel som bare gjelder det grafiske uttrykket:

$$\{sk\} \rightarrow \langle sk \rangle _V \in \{i, y\}, \langle skj \rangle _V \notin \{i, y\}$$

Vokalmengden som styrer reglene er forskjellig for den fonologiske regelen og den grafonomiske regelen.¹ I parentes kan det bemerkes at i nynorsk representerer det grafiske uttrykket de dypfonologiske formene i dette leksemet: <skyte>, <skyt>, <skaut>, <skote>

Også korrespondanser av typen <v>=/f/, <g>=/k/ og =/p/, <d>=/t/, altså der et ustemt konsonantfonem lisensierer et grafem med stemt primærfonem, kan forklares med dybdebegrepet.

$$\begin{aligned} \langle stiv \rangle &= /sti:v/ \sim \langle stivt \rangle = /sti:ft/ \\ \langle velge \rangle &= /velge/ \sim \langle valgt \rangle = /valkt/ \\ \langle redd \rangle &= /red/ \sim \langle redsel \rangle = /rets/ \\ \langle xenofob \rangle &= /seno^1fo:b/ \sim \langle xenofobt \rangle = /seno^1fo:pt/ \end{aligned}$$

Dette er vekslinger der konsonantens stemthet assimileres bort i kontakt med en ustemt konsonant, /s/ og /t/ i disse eksemplene, og for at dette skal være en morfofonemisk veksling, må det være en fonemisk opposisjon mellom stemt og

¹En annen mulig analyse er at den underliggende preteritumsformen er {skjø:t}, noe som vil gjøre dypstrukturen mindre regelmessig, men vil generere riktige grafonomiske og fonologiske overflateformer.

ustemt konsonant. En herding av /n/, /r/ eller /l/ i disse omgivelsene skaper ingen morfofonemisk veksling fordi stemthet her ikke har distinktiv funksjon og dermed ikke gir kommutasjon. Det er to konsonantpar til som er skilt bare ved stemthet, og det er /d/ : /t/ og /j/ : /ç/. Stavemåten for <torde> : <tort> indikerer at vi har en veksling av denne typen i dette bøyningensmønsteret, men jeg tror moderne SØ har <torde> = ²tu:te/ eller ²to:te/, slik at noen stemthetsveksling ikke finner sted. Andre par har jeg ikke greid å finne.¹ Siden /j/ har begrenset distribusjon og ikke forekommer morfemfinalt, finnes heller ikke bøyde eller avledede former der dette fonemet kommer i kontakt med ustemte konsonanter. Faktisk er det et argument for diftonganalysen av /æj/ og /øj/ at ord med disse diftongene i utlyd ikke får veksling med /ç/ i de aktuelle omgivelsene:

<grei> = /græj/ ~ <greit> = */græçt/

<høy> = /høj/ ~ <høyt> = */høçt/

Ordformene på s.99 har følgende dype former, og det er disse dype enhetene som er representert ved sine primærgrafemer i det grafonomiske uttrykket:

{stiv} ~ {stive} ~ {stivt}

{velg} ~ {velge} ~ {valgt}

{redd} ~ {redde} ~ {redsel}

{xenofob} ~ {xenofobe} ~ {xenofobt} ~ {xenofobi}

De korrekte fonologiske overflateformene blir generert av en enkel regel:

{v} → /f/ _K ∈ {/s/, /t/}, /v/ _V, /v/ _#

{g} → /k/ _K ∈ {/s/, /t/}, /g/ _V, /v/ _#

{d} → /t/ _K ∈ {/s/, /t/}, /d/ _V, /v/ _#

{b} → /p/ _K ∈ {/s/, /t/}, /b/ _V, /v/ _#

Formulert ved hjelp av distinktive trekk ville regelen bli enklere, men kanskje noe vanskeligere å lese. Jeg ønsker ikke å utvikle noen notasjon for dette her, da

¹Men *verdi* ~ *verdt* er en kandidat.

disse reglene har mer med fonologi enn grafonomi å gjøre og dermed faller utenfor oppgavens rammer. Regelen er dessuten uspesifisert for fonologiske omgivelser som ikke er diskutert over.

Også vekslingen mellom /nd/ og /n/ som lisensieringer for <nd> kan diskuteres i et dybdeperspektiv, for eksempel:

$$\{rund\}, a = \langle rund \rangle = /run/$$
$$\{runde\}, a = \langle runde \rangle = /rune/$$
$$\{runde\}, s = \langle runde \rangle = /runde/$$
$$\{runde\}, v = \langle runde \rangle = /rune/$$

En regel basert på utelukkende G/F-kontekst ville her ikke kunne generere de ønskede overflateformene:

$$\{nd\} \rightarrow /n/ _ \#, /nd/ _ V$$

I slike tilfeller må morfologiske eller leksikalske variabler inngå som parametre i regelen.

4.5.7 Homonymi

Et mye brukt argument for stavemåter som avviker fra det fonemografiske prinsipp, er argumentet om at rettskrivningen bør skille mellom homofonpar for å lette lesingen. Det hevdes altså at skriptet trenger leksikalske eller morfologiske markører for å sikre at korrekt korrespondanse i det grafonomiske tegnet etableres i leseretningen.

Vi kan gruppere ordpar (eller tripler, etc.) med homonyme relasjoner i ulike grupper etter forholdet mellom skrift og tale. Ettersom prosodemene tonem og trykk, og i en del sammenhenger kvantitet, ikke har systematisk dekning i bokmål, vil minimale par som skilles kun ved et prosodem, ofte få identiske skriftlige uttrykk dersom rent fonemografiske prinsipper legges til grunn for stavemåten. Jeg kaller slike minimale par *segmentalhomofoner* og behandler dem som en egen gruppe av homofoner under.

Homografene kan deles i tre grupper:

1. Homografer som også er homofoner, f.eks.
 - <rom>=/rom/="værelse" : "brennevin"
 - <verken>=²værken/="ingen av to" : "stoff av ull og bomull"
 - <for>=/får/="foran" : "årsakskonjunksjon"
 - <får>=/få:r/="mottar" : "sau"
2. Homografer som er segmentalhomofoner, altså homofoner om man ser bort fra fonologiske distinksjoner som har ingen eller delvis dekning i bokmål. I bokmål gjelder dette særlig tonem og trykk, men det forekommer kvantitetseksempler også:

- a. tonem: <milen>=¹mi:len/="av *mil*" : ²mi:len/="av *mile*"
- b. kvantitet: <spilte>=²spilte/="av *spille*" : ²spi:lte/="av *spile*"
- c. trykk: <kurer>=²ku:rer/="sykdomsbehandlinger" : /
ku¹re:r/="sendebud"

Alle eksemplene følger de grunnleggende grafonomiske reglene i bokmål. De er alle lineære en-til-en-representasjoner. I kvantitetseksempelet markeres den korte vokalen av dobbeltkonsonanten <lt>, mens den lange vokalen står i en overlang stavelse der den etterfølgende konsonantsekvens ikke kan representeres på noe annet vis enn ved dobbeltkonsonanten <lt>. I trykkeksemplet oppstår i tillegg en solidarisk kvantitetsdistinksjon,¹ der den aksentuerte stavelsens lengde realiseres ved lang vokal i begge tilfellene. I begge tilfellene er den lange vokalen markert med etterfølgende enkeltkonsonant <r>, mens den korte vokalen i den korte stavelsen også etterfølges av en enkeltkonsonant <r>. Begge deler er systemrett.

3. Homografer som ikke er homofoner, f.eks.

- <vannet>_s =¹vane/ : <vannet>_v =²vanet/

¹Se om overlange stavelser og solidariske prosodemer i kapittel 3.1.1 om fonemsystemet i SØ.

- <sort>=/soʔ/="svart" : <sort>=/såʔ/="slag"

Likeledes kan homofonene deles inn i fire grupper:

1. Homofoner som også er homografer, f.eks.

- a. <rom>=/rom/="værelse" : "sukkerbrennevin"
- b. <verken>=/ʔværken/="nektende konjunksjon"¹ : "type bomullsstoff"
- c. <fôr>=/fo:r/="mat" : "tøylag"
- d. <får>=/få:r/="sau" : "mottar"

2. Segmentalhomofoner som er homografer, f.eks.

- a. <tanken>=/¹taŋken/ av *tank* : /²taŋken/ av *tanke*
- b. <spilte>=/²spilte/ av *spille* : /²spi:lte/ av *spile*
- c. <kurer>=/²ku:rer/="botemidler" : /ku¹re:r/="sendebud"²

3. Homofoner som ikke er homografer, f.eks.

- a. /man/=<man>="en" : <mann>="voksen person av hannkjønn"
- b. /²jo:re/=<gjorde> av *gjøre* : <ljore>="røykhull"
- c. /²lå:ve/=<love>="gi løfte" : <låve>="lagerhus på gård"

4. Segmentalhomofoner som ikke er homografer, f.eks.

- a. /¹bøner/=<bønder> av *bonde* : /²bøner/=<bønner> av *bønne*
- b. /fu:lt/=<fult> av *ful* : /fult/=<fullt> av *full*
- c. /vi:lt/=<hvilt> av *hvile* : /vilt/=<vilt> av *vill*

¹Begge uttaleordbøkene jeg har brukt, har tonem 2 i dette ordet. Min intuisjon sier imidlertid at det også er vanlig å uttale dette ordet med tonem 1, i hvert fall dersom det ikke har emfatisk trykk. Ordet vil relativt sjelden være aksentuert i ordinære ytringer, og det vil dermed sjelden ha noen tonemrealisering i det hele tatt. Det virker dermed naturlig at tonem 1, som det synkront umarkerte tonemet, vil kunne spre seg i dette ordet.

² /²ku:rer/="botemidler" : /ku¹re:r/="sendebud" skilles primært ved trykkplassering, men ved aksentuering får også stavelsen lengde, og kvantitet vil dermed bli et subsidiært skille ved minimale par av denne typen.

- d. /²ale/= <alle>="hver og én" : /a¹le:/= <allé>¹="vei med trær på begge sider"

I 4b og 4c ser vi at kvantitetsdistinksjonen kan representeres både ved konsonantreduplikasjon og på andre måter.²

Det er viktig å merke seg at bokmål faktisk har et ganske stort antall homografer, tilsynelatende uten at dette virker særlig negativt inn på leseforståelse eller lesehastighet.³ I norsk, og i mye større grad i en del andre språk, for eksempel kinesisk, finnes dessuten et ikke uvesentlig antall homofoner. Det at disse homofonene finnes, uten at språket har utviklet nye virkemidler for å bli kvitt dem, tyder på at kommunikasjonen godt kan tåle et visst antall homofoner. Det hevdes⁴ at muntlig kommunikasjon har mange kompenserende virkemidler som skrift mangler, og at dette er grunnen til at muntlig kommunikasjon kan tåle homonymi, mens skrift, som mangler disse virkemidlene, trenger homografi-avvergende markører for å skille mellom homofoner. Jeg ønsker å knytte noen kommentarer til denne argumentasjonen:

1. Finnes det undersøkelser som underbygger påstanden om at muntlig kommunikasjon gjør bruk av eller er avhengig av homonymiavklarende virkemidler i særlig grad?
2. Mye muntlig kommunikasjon foregår i dag via teleteknologi, særlig telefon. Her er mange slike virkemidler fjernet; hemmes kommunikasjonen av dette?
3. Det er et faktum at svært mange av homofonene tilhører enten ulike grammatikalske kategorier eller ulike semantiske felt eller begge deler, og

¹ Ifølge Vinje 2004 kan aksenten utelates, men BMOB nevner ikke dette. Det er også mitt inntrykk at det er mest vanlig å bruke den. Imidlertid skiller ikke noen av de korpus jeg har brukt, mellom <é> og <e>, så jeg har ikke noe belegg for dette inntrykket.

²Se om dekning av kvantitet og konsonantreduplikasjon i 4.5.5.

³Jeg kjenner ikke til undersøkelser som underbygger dette inntrykket, og det er også en egenskap ved skriptet som det er vanskelig å gjøre kontrollerte undersøkelser om.

⁴For eksempel i Svennevig, Sandvik og Vagle 1995.

at det derfor er svært sjelden at det er reell fare for misforståelser i en gitt kommunikasjonssituasjon, enten den er muntlig eller skriftlig. Dog må det kunne antas at *lesehastigheten* vil kunne hemmes noe i tilfeller der de grammatikalske og/eller semantiske omgivelsene ennå ikke er klarlagt ved lesing av et ord som inngår i et homonymi-par.

Det er fire nivåer av homonymi:

1. Homonymer som har ulike rotmorfemer, men i samme grammatikalske kategori og samme bøyning, for eksempel:
 - <fôr>=/fo:r/="dyremat" : <fôr>=/fo:r/="tøylag"
2. Homonymer som har ulike grammatikalske kategori, men kanskje samme rotmorfem, for eksempel:
 - <baker>,s=/²ba:ker/ : <baker>,v=/²ba:ker/
3. Homonymer som har samme grammatikalske kategori, og kanskje samme rotmorfem, men ulik bøyning (og ulike leksem), for eksempel:
 - <finner>,s=/²fīner/="en som har funnet" :
 - <finner>,s=/²fīner/="personer fra Finland"
4. Homonymer som er samme leksem, men ulike bøyninger, for eksempel:
 - <hus>=/hu:s/="bygning" : <hus>=/hu:s/="bygninger"

4.5.8 Digrafer og funksjonelle enheter

Digrafer, trigrafer og mer generelt polygrafer brukes i litteraturen om mer eller mindre faste sekvenser av grafemer med mer eller mindre faste fonem-korrespondanser. Det finnes imidlertid ikke noen entydig definisjon som det hersker enighet om, og noen bruker termen i en så vid forstand at den mister en reell verdi.¹ <str> er for eksempel en relativt frekvent grafemsekvens i bokmål, og den er entydig og representerer fonemsekvensen /str/. Denne fonemsekvensen følger imidlertid også entydig fra linearitetsprinsippet i kombinasjon med disse grafemenes primærfonemer, og vi ville få et svært stort antall polygrafer i bokmål om vi opererte med en definisjon som inkluderte <str>.

¹For eksempel Allén 1971:45-46.

<skj> er imidlertid en relativt frekvent grafemsekvens som innenfor morfemgrenser er entydig og som *ikke* følger av enkeltgrafemenes primærfonem i kombinasjon med linearitetsprinsippet. <skj> er dermed en prototyp av en polygraf. Om vi bruker denne prototypiske definisjonen, får vi et vesentlig lavere antall polygrafer i bokmål, kanskje 7:

<skj>=/ʃ/	<skje>=/ʃe:/
<sj>=/ʃ/	<sjø>=/ʃø:/
<kj>=/ç/	<kje>=/çe:/
<rs>=/ʃ/	<særs>=/sæ:ʃ/
<rt>=/t/	<art>=/a:t/
<rn>=/ɳ/	<ørn>=/ø:ɳ/
<ng>=/ŋ/	<eng>=/eŋ/

Noen grafemsekvenser har to frekvente representasjoner i nøytrale kontekster, der den ene følger av linearitetsprinsippet og den andre ikke. I bokmål er dette

<tj>=/ç/, /tj/	<tjene>=/2tje:ne/	<tjern>=/çæ:ɳ/
<rd>=/d/, /rd/	<verdig>=/ ² værði/	<ferdig>=/ ² fæ:ɖi/

Disse grafemsekvensene kan sies å opptre som digrafer der de ikke følger linearitetsprinsippet og som grafemsekvenser ellers.

Visse grafemsekvenser har som primærfonem det ene grafemets primærfonem, mens det andre grafemet er "stunt". Det andre grafemet har fonologisk uttrykk null i denne posisjonen. Dette gjelder i bokmål:

<gj>=/j/	<gjø>=/jø:/
<hj>=/j/	<hjem>=/jem/
<hv>=/v/	<hva>=/va:/
<dt>=/t/	<godt>=/gåt/

Disse grafemsekvensene kan tolkes som digrafer, eller de kan tolkes som grafemsekvenser der det ene grafemet har nulluttale i visse omgivelser:

<g>=/Ø/|_j

<h>=/Ø/|_j

<h>=/Ø/|_v

<d>=/Ø/|_t

Primærfonemet til disse digrafene kan også forklares ut fra fonotaktiske begrensninger i SØ. Gitt disse fonotaktiske rammene kan ikke linearitetsprinsippet gjelde for disse grafemsekvensene. Flere digrafer har den samme egenskapen, men avhengig av kontekst. Konteksten kan skrive seg fra ulike lingvistiske nivåer:

<lj>=/j/ | #_ <ljå>=/jå:/ <detalj>=/de¹talj/

<ld>=/l/ | _+ <mild>=/mil/ <Hilde>=/²hilde/

<nd>=/n/ | _+ <runde>,v=²rune/ <runde>,s=²runde/

Vekslingen mellom verbformen og substantivformen i det siste eksemplet kan kanskje forklares ved å velge <rund> som stamme i verbformen. Da blir <nd>=/n/ ordfinalt og morfemfinalt foran bøyningsendelsen -e for infinitiv. I substantivet står <nd> ikke ved noen morfemgrense, noe som gir <nd>=/nd/. Denne forklaringen forklarer lignende vekslinger i <onde>,a=²one/ ~ <onde>,s=²onde/, men den blir for enkel til å forklare mangel på veksling i <avbilde>=¹a:vilde/ ~ <avbild>=¹a:vild/.

I <ki>=/çi/, <gi>=/ji/ og tilsvarende sekvenser med <y> og <ei> representerer digrafene en fonem*sekvens*. Eventuelt kan disse sekvensene analyseres som enkeltgrafemer <k>, <g> som har særlige egenskaper i visse omgivelser, nemlig | #_V ∈ {i, y, ei}, mens vokalgrafemene følger linearitetsprinsippet. Alle disse digrafene har dessuten unntak som bare kan forklares leksikalsk.

<kikk>=/çik/	<kivi>= ¹ ki:vi/ <kick>=/kik/ <kis>=/ki:s/ <kiosk>=/çåsk/ ¹
<gikk>=/jik/	<gir>=/gi:r/ <gigant>=/gi ¹ gant/ <giro>= ¹ ʃi:ro/ <gin>=[dʒin], /ʃin/ ²
<kyss>=/çys/	<kyrillisk>=/ky ¹ rilisk/ <kykeliky>=/kykeli ¹ ky:/
<gys>=/jy:s/	<gylf>=/gylf/ <gym>=/gym/
<keitet>= ² çæjtet/	<keiser>= ² kæjser/
<geit>=/jæjt/	<Geir>=/gæjr/ <geigerteller>= ¹ gæjgeteler/

I en særstilling blant digrafene er de som representerer diftonger, i og med at diftonger realiseres som en glidning i artikulasjon. At diftonger lisensierer digrafer, ligger derfor ofte nærmere linearitetsprinsippet ved at det første grafemets primærfonem ligger nær diftongglidningens startpunkt, og det andre grafemets primærfonem ligger nær diftongglidningens sluttunkt, f.eks. <ai>=/aj/. Imidlertid er likheten ikke så dominerende i diftongene <ei>=/æj/ eller <oi>=/åj/³.

I <eg>=/æj/ ligger grafemenes primærfonemer enda lenger fra diftongens endepunkter, og dessuten opptrer <eg>=/eg/ også som en ordinær grafemsekvens, f.eks. <treg>=/tre:g/ ~ <tregt>=/tre:kt/.

¹I <kiosk> representerer grafemsekvensen <ki> enkeltfonemet /ç/ og kan analyseres som en digraf.

²Jeg tror en uttale med engelsk [dʒ] er vanligst her, noe som gjør ordet til et fremmedord, ettersom denne fonen ikke er del av SØ fonologi. En tilpasset uttale med /ʃ/ er nok heller ikke uvanlig.

³<oi>=/åj/ ville bli regelmessig om man inkluderte <i> i de etterfølgende omgivelsene som gir <o>→/å/.

Vi begynte diskusjonen av polygrafer med å fastslå at de opptrer innenfor morfemet. Dersom en morfemgrense opptrer innenfor en grafemsekvens som er en potensiell polygraf, vil grafemsekvensen ofte følge linearitetsprinsippet:

<brekkjern>=¹brekjæ:ŋ/

<huljern>=²hu:ljæ:ŋ/

Lignende eksempler kan finnes eller konstrueres for de fleste grafemsekvenser som er potensielle polygrafer, men ikke for de polygrafer som har som primærfonem enkeltfonem som er synkrone assimilasjonsprodukter av enkeltgrafemenes primærfonem, for eksempel <rs>=/ʃ/. Blant de tusen mest frekvente ordene i Oslo-korpuset finnes 28 ordformer med grafemsekvensen <rs>, og alle representerer /ʃ/, også der grafemsekvensen krysser en morfemgrense. /r/+/s/ → /ʃ/ er en produktiv assimilasjon som også virker over ordgrenser. Vi har valgt å kalle <rs> en digraf for /ʃ/, men i lys av dette er det også mulig å tolke <rs> som en ordinær grafemsekvens som følger linearitetsprinsippet og representerer dypstrukturen {rs}, og at det er synkrone fonologiske prosesser som sørger for at det fonologiske overflateuttrykket blir /ʃ/.

Digrafer kan fungere som funksjonelle enheter¹ eller som en dobbel enhet. I ord som <sart>=/sa:t/ og <Lars>=/la:ʃ/ fungerer digrafene som funksjonelle enheter i og med at de ikke markerer for lang vokal. I ord som <tang>=/taŋ/ og <hesj>=/heʃ/ fungerer digrafene som dobbeltenheter, altså som grafemsekvenser, i og med at de markerer for kort vokal. Alle eksemplene over er røtter; i bøyde former som <rant>=/ra:t/ og <fars>=/fa:ʃ/ er stavelsen kanskje overlang, som i <bilt>=/bi:lt/, og vi kan derfor ikke bruke de bøyde formene til å vise noe om funksjonelle enheter.

<ng> opptrer aldri som en funksjonell enhet; den følger alltid etter kort vokal. Imidlertid er ikke dette nødvendigvis et uttrykk for noen egenskap ved digrafen, i

¹Venezky 1970: *functional entities*.

og med at fonotaktiske regler i SØ umuliggjør stavelser med [V:ɲ]. Om import eller lydendringer introduserte [V:ɲ] som en mulig stavelse i SØ, ville den trolig lisensiere <Vng>. Digrafer som <dt>=/t/, <nd>=/n/ og <ld>=/l/, derimot, opptrer i fonologiske omgivelser der det er mulig med både lange og korte vokaler: [Vt] og [V:t], [Vn] og [V:n], [Vl] og [V:l]. Ettersom disse digrafene alltid markerer for kort vokal, opptrer de alltid som dobbeltenheter.

<rs>, <rt> og <rn> ser ut til å kunne fungere på begge måter, som vi ser av eksemplene:

<barn>=/ba:ɲ/, <morn>=/mɑɲ/

<Lars>=/la:ʃ/, <vers>=/væʃ/

<sart>=/sa:t/, <fort>=/foʈ/

<rd> virker som en funksjonell enhet der den fungerer som en digraf, altså ved <rd>=/r/ og <rd>=/d/, men ikke ellers:

<ferdig>=/²fæ:di/, <hard>=/ha:r/

<ferd>=/færd/

Dessuten går det fram at det ikke er slik at alle digrafene av typen der det ene grafemet har nulluttrykk (se over), har de samme egenskapene. <rd>=/r/ er en funksjonell enhet, mens <nd>=/n/ ikke er det.

<sj> synes å virke som en funksjonell enhet i flerstavellesord med trykket på siste stavelse, men som en dobbeltenhet i enstavellesord eller der trykket er på første stavelse:

<æsj>=/æʃ/, <hasj>=/haʃ/, <pysj>=/pyʃ/, <hesje>=/²heʃe/

<fromasj>=/fro¹ma:ʃ/, <pastisj>=/pa¹sti:ʃ/

/ç/ har sterkt begrenset distribusjon, men der den forekommer stavelsesfinalt, i *bikkje* og *krykkje*, ser det ut til at normen oppfatter digrafen <kj> som en

funksjonell enhet, ettersom det er brukt en trigraf for å sikre kort uttale i disse ordene. Ingen andre digrafer tar i bruk dette potensielle virkemiddelet for å markere vokalkvantitet:

/heŋ/ = * <henng>

/mån/ = * <morrn>

/foŋ/ = * <forrt>

/væŋ/ = * <verrs>

Vi ser av punktene over at digrafenes opptreden som funksjonelle enheter kanskje i større grad er en funksjon av fonotaktiske begrensninger i de aktuelle stavelsene enn en iboende egenskaper ved digrafene.

Grafemene <m> og <j> er også tvetydige med hensyn til om de er funksjonelle enheter eller dobbeltenheter:

<dåm> = /då:m/, <dom> = /dåm/, <dom> = /do:m/, <dommer> = /²dåmer/
<never> = /¹n ever/, <lever> = /¹lever/, <le:ver> = /¹le:ver/,

4.5.9 Grafotaks

Grafotaksen kan deles inn i to kategorier:

1. Grafotaktiske regler i rettskrivningsnormen som er paralleller til fonotaktiske regler i det talte språket. For eksempel:

- a. Bare <s> eller <Ø> ordfrontalt foran <pr>
- b. Ikke <ng> ordfrontalt.
- c. <h> bare morfemfrontalt.

Med fonemklammer rundt primærfonemet i stedet for grafemklammene ville alle disse vært gyldige fonotaktiske regler.

2. Grafotaktiske regler i rettskrivningsnormen som ikke reflekterer fonotaktiske regler i det talte språket. For eksempel:

a. Forenkling av trippelkonsonant:

$$\langle K_1 K_1 K_1 \rangle \rightarrow \langle K_1 K_1 \rangle, \langle K_1 K_1 - K_1 \rangle$$

b. Forenkling av dobbel <m> ord- eller morfemfinalt:

$$\neg \langle mm \rangle \mid _ \# , _ + \text{ (eventuelt } \langle mm \rangle \# \rightarrow \langle m \rangle \text{)}$$

c. Ikke dobbel <v> unntatt over morfemgrense:

$$\neg \langle vv \rangle \mid + _ _ + \text{ (innenfor ett morfem)}$$

Konsekvensen av b. og c. er at kort vokal foran disse grafemenes primærfonemer ikke markeres i skriptet. Derimot er tilsvarende regel for <j> relatert til at grafemets primærfonem /j/ ikke opptrer stavelsesfinalt og derfor ikke kan redupliseres.

Distribusjonen av <hv> er en grafotaktisk regel som ikke har noen parallell i fonotaktiske regler. /v/ har full distribusjon, så distribusjonen av /v/ er ikke det som utelukker digrafen i grafonomiske uttrykk som

$$/na:v/ = * \langle nahv \rangle$$

$$/^2ga:ve/ = * \langle gahve \rangle$$

$$/^1never/ = * \langle nehver \rangle$$

Imidlertid begrenser distribusjonen av /h/ distribusjonen av grafemsekvensen <hv>=/hv/, som er null.

De grafotaktiske begrensninger som regulerer distribusjonen av digrafene med <j>, derimot, er direkte paralleller til de fonotaktiske begrensninger som regulerer distribusjonen av /j/. I og med at /j/ ikke kan opptre finalt, kan heller ikke disse digrafene opptre finalt:

$$* \langle hagj \rangle, * \langle halj \rangle, * \langle hahj \rangle$$

Grafemsekvensen <lj>=/lj/ har ikke slike begrensninger, jf. <detalj>.

Mange av skriptets konvensjoner er i liten grad relatert til fonotaks eller andre egenskaper ved den fonologiske strukturen. For eksempel er markeringen av kort vokal ved dobbelkonsonant en tilfeldig grafotaktisk egenskap som ikke har noen relasjon til det fonologiske system, ettersom vi har analysert distinktiv kvantitet som en egenskap ved vokalen. I et fonemografisk perspektiv kunne like gjerne kvantitet markeres ved dobbeltvokal, noe som også ville være en mer presis gjengivelse av både de fonemiske og de fonetiske forhold (Vanvik 1979):

/ha:t/=*<haat>, /hat/=*<hat>

Nederlandsk bruker blant annet denne konvensjonen for å markere vokallengde, og Usenet-korpuset viser at den også har en viss aksept i norsk skrift – i ett leksem. <kuul>, <kuul> og <kuule> har til sammen 518 forekomster, noe som utgjør omkring 7% av bruken av dette leksemet.¹

4.5.10 Gjennomsiktighet

Et mye brukt skille (f.eks. Danielsson 2003) er *gjennomsiktighet* kontra *ugjennomsiktighet*². Termene kan brukes både på ordnivå og på skriptnivå. På ordnivå reflekterer gjennomsiktighet i hvilken grad et ords uttale kan utledes av stavemåten, mens på skriptnivå betegner det hvilken tendens til gjennomsiktighet på ordnivå det er i skriptet. Danielsson klassifiserer gjennomsiktighet for enkeltord på tre nivåer:³

1. Fullstendig gjennomsiktige ord, der det eksisterer en en-til-en-korrespondanse mellom grafem og fonem, f.eks. <ben>⁴ eller <bli>.

¹ Dette er også flere forekomster enn den engelske stavemåten <cool>.

² Den engelske termene er *transparent* / *transparency* og *opaque* / *opacity*.

³ Danielsson skriver om svensk rettskrivning, men de teoretiske betraktningene er allmenngyldige. Svensk og bokmål rettskrivning har også mange parallelle egenskaper, skjønt de er ikke identiske.

⁴ Dette er også Danielssons eksempler, og de er enda mer entydige for svensk enn for bokmål. For bokmål er selvfølgelig <ben> kun entydig dersom man lar den korrespondere med uttalen /be:n/.

2. Regelmessige ord, men gjennomsliktige bare innenfor G/F-kontekst, f.eks.

- a. Trykksterk kort vokal etterfulgt av geminert konsonantgrafem eller to ulike konsonantgrafer, f.eks. <pappa>, <hest>.
- b. Trykksvak kort vokal etterfulgt av enkelt konsonantgrafem, f.eks. <tusen>, <bilen>.
- c. Digrafer, f.eks. <kjele>, <sang>.

3. Irregulære ord, f.eks. <og>, <seksten>

Det er flere aspekter denne definisjonen av gjennomsliktighet ikke tar hensyn til. For det første omtaler den bare leseretningen, om korrespondansen er entydig i leseretningen. Korrespondansene kan være entydige i leseretningen og ikke i skriveretningen, og omvendt:

<ljore> → /²jo:re/ er entydig i leseretningen.

/²jo:re/ → <ljore>, <gjorde>, <jorde> er ikke entydig i skriveretningen

<egne> → /²e:gne/, /²æjne/ er ikke entydig i leseretningen

/²e:gne/ → er entydig i skriveretningen

ljore er altså entydig i leseretningen, men ikke i skriveretningen, mens *egne* er entydig i skriveretningen, men ikke i leseretningen. Disse egenskapene er direkte funksjoner av forholdet mellom homofoni og homografi. Entydighet i leseretningen men ikke omvendt, altså det Danielsson kaller gjennomsliktighet, vil ofte være relatert til fonotaktiske begrensninger. Slik er <hval> → /va:l/ entydig ettersom /hva:l/ ikke er en lovlig stavelse i SØ, mens /va:l/ ikke er entydig fordi /v/ kan lisensiere flere grafemer.

Entydighet i skriveretningen men ikke omvendt, derimot, vil ofte være relatert til dekningsgrad. Slik vil /¹baner/ være entydig, men <baner> kan representere ordformer med begge tonem, ettersom tonem ikke er dekket i bokmål.

Det finnes også svakere former av gjennomsiktighet, der G/F-kontekst lisensierer andre grafem enn primærgrafemet, for eksempel /å/ → <o> | _<ng>. Dette er definitivt entydig i bokmål; alle ord på /ång/# og alle ord på <ong># i *Bokmålsordboka* følger denne regelen, men det er uvisst om det tilfredsstillende kriteriene for Danielssons gjennomsiktighetsbegrep. Vi kan kalle denne typen korrespondanser for *regelmessige* i bokmål.

Danielssons gjennomsiktighet er beslektet med termen *ortofoni*, men dette er en term som er kjent, og som har et heller vagt begrepsinnhold, og som vi derfor ikke bruker.

5 Bokmål som fonemografisk skrift

5.1 Vokaler

5.1.1 Vokalfonemer

fonem	grafem	fremmed	lange	korte
i	i e		bi il bil de	bitt
y	y		by yr syl	lytt
e	e æ		se ed ler sæd svæve træ	fett trædd
ø	ø y	u	mø ør skjør	bløtt syttten sytti pub
æ	e æ	a	er sær	tredve verre æsj færre bag
a	a		da ad fat	satt
u	u		bu ul skjul	skuff,v
o	o u		bo ord mor	loff plomme lompe skuff,s bunke lunge kum
å	o å		oktober over gå år går	kott flott plomme hålke flått fråtse

Vi ser at i hjemlige ord kan

- /i/ lisensiere <i> eller <e>, men <e> bare i ett leksem,
- /e/ lisensiere <e> eller <æ>,
- /æ/ lisensiere <æ> eller <e>,
- /o/ lisensiere <o> eller <u>,
- /å/ lisensiere <å> eller <o>,
- /ø/ lisensiere <ø> eller <y>, men <y> bare i én stamme.

/a/, /i/, /y/, /u/ lisensierer kun sine primærgrafem, med unntak av i visse fremmedord. Eksemplene er imidlertid importord som er fullstendig assimilert

morfologisk. *Pub* er også assimilert fonologisk, mens *bag* bryter med distribusjonsbegrensningene for /æ/.

En del ord med <o> har fakultativ veksling mellom /o/ og /å/, for eksempel <plomme>=/plome/, /plåme/. Dette er også tilfellet i en del ord med <o> i trykksvak førstestavelse, for eksempel

<korrespondent>=/kårespån¹dent/, /korespon¹dent/

Distinksjonen /o/ : /å/ kan kanskje sies å ha uklar fonemisk status i en del leksikalske og/eller fonologiske kontekster,¹ og i hele det bakre vokalområdet er det mye variasjon i uttalen og vanskelig å finne minimale tripletter mellom korte vokaler.

<e> og <æ> representerer /e/ og /æ/, skjønt ikke med noe mønster som kan defineres ut fra G/F-kontekst. I tillegg til rammene som /æ/ sin begrensede distribusjon legger, er det i stor grad leksikalske (etymologiske) prinsipper som styrer fordelingen av <e> og <æ>.

<y>=/ø/ bare i to leksem med samme rot.

5.1.2 Vokalgrafemer

grafem	fonem	fremmed	eksempler lange	eksempler korte
a	a	æ æj	sa ak bak bag tape	ball man bag
e	e æ i Ø	a Ø	se ser allé lampe bety er de elleve lime tape	tredje redd men tredve entré
i	i j <i>diftong</i>	i aj	si is spist ion lime hai hei hoie huie	titt til
o	o å		lo or kost over oktober	kost kost oktober for

¹Haslev 1985.

u	u o <i>diftong</i>	ø	du ul gul sau	puff sum hun dum pukk pub
y	y ø i <i>diftong</i>		sy yt syl Dagny sexy høy høyt	pytt synd syttten sytti
æ	e æ		svæve sæd lær svært fælt	trædd færre
ø	ø		øm blø søt	smør øl bør
å	å		lå ål går	flått

Vi ser at i hjemlige ord kan

- <e> representere /e/ og /æ/, og dessuten /ə/ om vi regner schwa som eget fonem,
- <o> representere /o/ og /å/,
- <u> representere /u/ og /o/,
- <y> representere /y/, /ø/ og /i/, men /ø/ bare i én stamme, og /i/ bare i trykksvak utlyd. Mye tyder på at opposisjonen mellom /y/ og /i/ er nøytralisert i denne posisjonen,
- <æ> representere /æ/ og /e/.

<a>, <i>, <ø> og <å> representerer som monoftonger kun sitt eget primærfonem. <i>, <y> og <u> inngår som sisteledd i diftongdиграfer. <a>, <e>, <o>, <u>, <ø> inngår som førsteledd i diftongdиграfer.

5.2 Diftonger

SØ har en rekke diftonger, én mot /u/ og resten mot /j/. Noen er sentrale elementer i norsk fonologi, mens andre er perifere, og bare til stede i et mindre antall ord. De sentrale diftongene er /æj/, /aj/, /øj/, /æu/, ifølge NUOV, mens /åj/, /oj/ og /uj/ er mer perifere. De to sistnevnte har ikke Vanvik en gang listet opp i oversikten over diftonger i forordet til *Norsk Uttaleordbok*, men de er brukt i transskripsjoner i ordlista.

Diftonger er et interessant område av bokmål, for her kan det argumenteres for at bokmål representerer allofoniske variasjoner i talen. En klassifisering av diftongene som glidninger fra en vokal til en av halv vokalerne [j] eller [w], slik Kristoffersen 2000 gjør, er ikke en urimelig tolkning rent fonetisk. Også Vanvik 1979 sier at fonetisk er det knapt forskjell på sluttpunktet i diftongen /æj/ og /j/, mens det blir verre å innta et tilsvarende standpunkt for /æu/, ettersom vi ikke har noen halv vokal [w] med fonemisk status ellers i SØ. Kristoffersen transskriberer riktignok med [w] i <duell>=[du¹wel], men Kristoffersens generative fonologi opererer med et annet fonembegrep, og innenfor et strukturelt paradigme er det ikke vanlig å regne med noe slikt fonem som /w/ i SØ.¹ I en strukturalistisk modell er det mest naturlig å definere diftonger som en glidning fra ett fonem til et annet, og om vi ikke regner med et fonem /w/ i SØ, blir det mest naturlig å definere diftongen som /æu/.

Når det gjelder diftongene som har glidning mot en fremre artikulasjon, er det minst to muligheter. Vanligst er det å definere diftonger som glidning fra et vokalfonem til et annet vokalfonem, og dette er også den tradisjonelle måten å gjøre det på for norsk, med de sentrale diftongene definert som /æi/², /øy/ og /ai/. Slik er det da også Vanvik definerer dem i *Norsk Uttaleordbok*. Vanvik diskuterer ikke de perifere diftongenes fonetiske egenskaper, men det virker mest sannsynlig at rundingen av startpunktet er overført til sluttpunktet deres, slik at de ut fra fonetiske egenskaper best kan klassiferes som /åy/, /oy/, /uy/.

Selv om diftongene nå er transskribert på en måte som er tilfredsstillende rent fonetisk, er det et spørsmål om det er formålstjenlig å skille mellom sluttpunktene /i/ og /y/ i en fonemisk transskripsjon. Det finnes ingen minimale par med kontrast mellom rundet og urundet sluttpunkt i diftongene, */Vi/ : /Vy/. Det eksisterer med andre ord ingen opposisjon mellom diftonger med ulikt fremre sluttpunkt. Det virker dermed som en bedre beskrivelse å si at rundingen er

¹ Vanvik 1979, Sivertsen 1967, Haslev 1985.

² Det har også vært vanlig å definere denne diftongen som /ei/, som er mer i tråd med skriftbildet, og sannsynligvis også med tidligere tiders uttale. I dagens SØ er det imidlertid liten tvil om at startpunktet for diftongen ligger nærmere /æ/ enn /e/ (Vanvik 1979).

nøytralisert og la den være et ikke-distinktivt allofonisk trekk ved diftongen: /VV_{fremre}/. Problemet består i så fall i hvilken vokal vi skal velge som sluttpunkt. Ettersom begge de aktuelle vokalene /i/ og /y/ har runding som ett av sine distinktive trekk, vil bruken av én av dem kunne virke misvisende, og det er i hvert fall ingen god beskrivelse uten en forklaring som viser at rundingen i denne posisjonen er nøytralisert. Ved å velge halvvokalen /j/ som sluttpunkt for diftongen, slik Kristoffersen gjør, riktignok i en generativ modell, unngår vi dette problemet, ettersom /j/ ikke har runding som ett av sine distinktive trekk.

fonem	grafem	fremmed	eksempel
æj	ei eg ek	a_e	ei hei sneip veg regne seksten tape
øj	øy		øy høy høyt
æu	au		au sau flaut
aj	ai aj	ay i_e	hai kai kaie kajakk bajas raje bayer ¹ lime
åj	oi		koie hoie ²
oj	oi oj		hoie jojo ³
uj	ui		huie

Vi ser at med denne beskrivelsen av diftongfonemene representerer <y> i <øy> den allofoniske rundingen av halvvokalen /j/, i motsetning til fraværet av runding i halvvokalen i <ai> og <ei>. Vi har dermed et subfonemisk aspekt ved bokmål i ord som *høy*, *bøyle*. I og med at /j/ i /åj/, /oj/, /uj/ ikke lisensierer <y> i ord som *oie*, *hoie*, *huie*, er ikke dette et gjennomført prinsipp for diftongene.

¹ *Hai* og *kai* fra nederlandsk *haai* og *kaai* (BMOB). *Kajakk* fra eskimo (BMOB). *Bajas* fra italiensk *bajazzo* (BMOB). *Bayer* fra tysk *Bayern* (BMOB). *Raje* av *rå* (RMOB).

² NUOV oppgir uttaler av *hoie* både med /åj/ og med /oj/. *Koie* av nedertysk *kôje*.

³ Vanvik transskriberer i NUOV dette ordet, som han for øvrig staver <yo-yo>, med fonetisk lang /j/ - altså en kort vokal foran /j/, noe han ellers ikke gjør, og noe som heller ikke er vanlig å gjøre for SØ. Det er fristende å tro at dette er en arbeidsulykke som er forårsaket av det grafiske uttrykket han bruker for ordet. Jeg transskriberer det med diftong: /²jojo/.

Om vi velger å beskrive diftongene som glidninger til vokalene /i/ eller /y/, får vi en fonemisk uregelmessighet, idet /y/ i de aktuelle ordene blir representert av <i>.

Uansett hvilken modell vi velger, finner vi en veksling i representeringen av sluttpunktet mellom <i>, <y>, <g> som er vanskelig å forklare ut fra et synkront perspektiv.

grafem	fonem	fremmed	eksempel
ai	aj		hai kaie
aj	aj		bajas raje
au	æu		au sau flaut
ay		aj	bayer
eg	æj		jeg negl
ek	æj		seksten sekstende
ei	æj		ei hei sneip
oi	åj oj		koie oie joik troika hoie hoie
oj	oj		jojo
ui	uj		huie
øy	øj		øy høy høyt

Konsonantgrafemene <j>, <g> og <k> kan kombineres henholdsvis med <a> og <o> og med <e> for å danne diftongdigrafer. <eg> er en etablert digraf for /æj/, særlig i frekvente ord. De personlige pronomenene *jeg*, *meg*, *deg*, *seg* trekker opp. <ek>=/æj/ bare i én stamme.

Mange av ordene med /aj/ eller /åj/ er importord, men de fleste er ikke fremmedord, for eksempel *hai*, *kai*, *bajas*, *koie*. Mange andre er fullstendig assimilert fonologisk, men har en semantisk karakter av fremmedartethet: *jojo*, *joik*, *boikott*, *troika*, *kajakk*.

5.3 Konsonanter

5.3.1 Konsonantfonemer

fonem	grafem	fremmed	eksempler lange	eksempler korte
p	p b		pal ap ape	oppe opp Ibsen obskur xenofobt

fonem	grafem	fremmed	eksempler lange	eksempler korte
t	t dt d		tå at ute	hatt godt endt redsel
ʈ	rt rdt rd lt t		sart fælt	ert hardt burde Arnt Ernst ?varmt
k	k g		ku ak ake	takk sagt viktigst
b	b p #s_		bo stab spe	labb
d	d t #s_		dø øde bod stå	ladd bed redde
ɖ	rd d		ferdig <er du> barndom	ferdig verdi
g	g k #s_		gå skog lage sko	egg egge
m	m		må lim	øm komme
n	n nd		nå lin syne	onn unna ond ende
ɳ	rn ren n rm rgen		barn morgen	læreren furten kanskje ?varmt
ŋ	ng g _n# n nt			sang sagn hank bronse entré interessant
l	l ld rl rel lv gl		le ul fæl farlig værelse fugl	øl ull til falle delje ild milde tolv halv sølv
[ʌ]	rl l		herlig mulig	
[ɾ]	l		fæl	delje
f	f v		få stivt	paff stift

fonem	grafem	fremmed	eksempler lange	eksempler korte
s	s ts ps x z c		si is oss tsar psyke xylofon zoologi celle	
ʃ	sj skj #_ sk rs #_ s s #_1 g #_ j ch tsj sgj rc sh		sjø sjy hasj skjell sky ski skøyte Lars særs svær Sveits stilig slå giro generell ajourføre tranchere tsjekker nysgjerrig ¹ shipping	hasj æsj vers Mercedes dash
ç ²	kj #_ kkje tj #_ k #_ i,y,ei		kjerne tjern tjei kei kiste kyss	bikkje krykkje
h	h		ha	
v	v hv Hu	w	vit leve verken hvit hvis Huitfelt wigwam	lever never lavvo
j	j gj hj lj g i		ja bejae gjø medgjørlig hjelp behjelpelig ljå ljore gi gyse geip ion	
r	r rd Rj Hr		rar mare hard borde Rjukan Hroar	for bør tør tørr irre

¹Både NROB og BMOB indikerer at *nysgjerrig* er importert av nedertysk *nisgirig*. I så fall kan ordet neppe sies å ha en synkron morfemgrense *{nyss#gjerrig}, selv om dette trolig er en populær folkeetymologi og også har støtte i kildespråket.

²Hos mange unge språkbrukere i blant annet Oslo er frikativene /ʃ/ og /ç/ falt sammen. Kanskje vil dette spre seg til standardtalemålet etter hvert, men det kan neppe regnes som del av SØ i 2005.

Når det gjelder korrespondansene <t>=/d/, <k>=/g/, <p>=/b/, må det knyttes en kommentar til dem. Disse korrespondansene er bare gyldige i posisjon etter <s> frontalt, og de skriver seg fra at fortis/lenis-distinksjonen er nøytralisert i denne posisjonen. Fonen som følger /s/ i ord som *spå*, *stå*, *sko*, kalles gjerne et arkifonem og er altså uspesifisert med hensyn til de trekk som normalt skiller plosivparene. Å hevde at disse fonene tilhører lenis-plosivene er dermed tvilsomt, men ikke mindre gyldig enn å hevde det motsatte.¹

Ellers er det påfallende å se hvor få konsonantfonemer som bare kan lisensiere ett grafemisk uttrykk. Dette gjelder bare /h/.

I rask tale vil /n/ gjerne assimileres til /m/ i ord som <brannbil>=¹brambi:l/ og <kronprins>=¹kromprins/, mens de fleste språkbrukere nok vil ha /n/ i disse ordene i saktere tale. Mange grammatiske "småord" har forskjell mellom sterke og svake former, der de svake formene brukes når ordet er uaksentuert i en ytring. Vanvik 1985 oppgir både sterke og svake former, for eksempel:

<av>=/a:v/, /a:/ | {sterk form}, <av>=/a/ | {svak form}
 <jeg>=/jæj/ | {sterk form}, <jeg>=/jæ/ | {svak form}²

I tilfeller av ord med flere uttalevarianter avhengig av talehastighet eller aksentuering er hovedregelen at bokmål representerer de sterke og de "sakte" formene.³

Det er imidlertid én type unntak, nemlig en del grammatiske "småord" med kort vokal som likevel skrives med enkelkonsonant; den korte vokalen er med andre

¹Det er vanlig hos barn som er i en fonemorientert fase i skriveutviklingen, (Hagtvatn 1996) at arkifonemet lisensierer primærgrafemet til den ustemte plosiven og gir uttrykk som <SGO> og <SDÅ>.

²At de svake formene har kort vokal i utlyd viser entydig at stavelsen er trykksvak.

³I engelsk er det mer vanlig å gjengi visse svake former, som <n't>=<not> og <I'm>=<I am>. I bokmål er dette i bruk kun i gjengivelse av direkte tale der forfatteren ønsker å indikere et stilnivå på talen, for eksempel <kan'ke>=<kan ikke>, der apostrofen representerer én eller flere eliderte bokstaver. Denne bruken er konvensjonell, men ikke normalisert, og vi regner den ikke som del av verken operativ eller kodifisert bokmålsnorm.

ord ikke markert, slik den er ellers i andre enstavelsesord. Krogsrud og Seip 1942 oppgir 30 av disse ordene, inkludert enkelte bundne røtter: *an, at, bør, den, din, en, et, for, gjen-, han, hen, hos, hun, hvis, kan, kun, man, men, min, mis-, mon, nok, når, sin, skal, til, tør, van-, vel, vil*, og bruker begrunnelsen at de oftest er trykklette og dermed at konsonanten oftest er kort. (Etter min analyse av fonologien i SØ i 3.1.1 ville jeg si at stavelsen er kort.)

Med unntak av *min, din* og *sin*, som har varianter med kort og lang vokal, har ikke disse ordene lang vokal i sine sterke former, så det er ikke sterke former som er representert, men svake. Når det gjelder leddene *gjen-, mis-* og *van*, som i *gjengi, mistanke* og *vantrives*, er vel forklaringen til Krogsrud og Seip om frekvente uaksentuerte posisjoner mindre presis.¹

Diskusjon om assimilasjonsprodukter reiser også flere problemstillinger. For det første blir produktene av historiske, avsluttede assimilasjonsprosesser ofte gjengitt fonemografisk. Dette gjelder mange arveord, men det er også et vanlig trekk ved en del importord fra klassiske språk, der assimilasjonen har funnet sted *før* importen til norsk:

<hente>=/²hente/ ← *heimta*
 <illegal>=/¹ilega:l/ ← *inlegal*
 <import>=/im¹påt/ ← *inport*
 <irregulær>=/¹iregulæ:r/ ← *inregulær*²

For det andre gjengis produktene av synkront aktive assimilasjonsprosesser gjerne morfemografisk, altså ved å holde morfemrepresentasjonen konstant uavhengig av fonologiske omgivelser:

<barndom>=/²ba:ŋdåm/
 <dersom>=/¹dæfåm/

¹Variasjonen i bruken av aksent i <en>:<én> og <ei>:<éi> kan også analyseres som en representasjon av svake og sterke former. Se kapittel 5.3.1.

²Dette er hypotetiske norske former, og ikke de historisk korrekte latinske formene.

Disse ordene blir sjelden eller aldri uttalt uten retroflekser (med /d/ og /s/) av SØ-talere, men assimilasjonene er produktive, og derfor er det de isolerte morfemenes fremlydsfonem som lisensierer grafemet. Det er først når assimilasjonen ikke lenger er produktiv, og uttalen eventuelt stivner med retrofleks uttale, at slike sammensetninger kan bli ugjennomsiktige og dermed lisensiere en mer fonemografisk uttale, som /^ldæʃām/=*<dæsjom>. For *barndom* kommer det tilleggsproblem at bokmål ikke har noen mekanisme for å representere stavelsesfrontal retrofleks dental.

5.3.2 Konsonantgrafemer

<i>grafem</i>	<i>fonem</i>	<i>fremmed</i>	<i>eksempler lange</i>	<i>eksempler korte</i>
b	b p		bo tube	labb Ibsen obskur
c	s	k ʃ tʃ ks	celle cash cover cup score cello	decennium back broccoli cappuccino accent
d	d t Ø ɖ		død døde stod god barndom	dødd redde onde godt redsel ild onde
f	f v		fy sjef sløyfe sfære	uff luffe
g	g k j ŋ Ø <i>diftong</i>		gave sag gi gyse gjøre mulig drog fugl regn døgn løgn negl jeg	legg legge sagt viktigst sagn signal
h	h Ø		hi hjul hval godhet	
j	j <i>diftong</i> Ø		jag jojo raje kajakk Rjukan	

<i>grafem</i>	<i>fonem</i>	<i>fremmed</i>	<i>eksempler lange</i>	<i>eksempler korte</i>
k	k ç <i>diftong</i>		kul kake kiste kyss seksten	hekk hekke
l	l Ø [ɾ] [ʌ]		le øl ljå fæl værelse mulig	ull tulle delje
m	m		må lim åme	øm komme
n	n ɲ ɳ		nå lin ane	inn unna tank bronse entré kanskje furten
p	p b Ø		på spe psyke	opp
q		k	qua Qatar	
r	r Ø		rar sure herlig	harr surre
s	s ʃ		si is lyse slå svær Sveits stilig	spiss passe
t	t d Ø		tut pute huset	smett kjette
v	v f Ø		viv sive lever stivt gav sølv	lever
w		v <i>diftong</i>	wigwam knowhow	
x		ks s	taxi xylofon	
z		s	zoologi quiz	

Også når vi ser på korrespondansen fra grafem til fonem, er det påfallende hvor få entydige korrespondanser som finnes, bare <m z q> representerer kun ett fonem, i tillegg til <r> og <h> som kan representere ett fonem pluss null.

<dg h j l p r t v> kan alle representere null.

I fremmedord som <design>=/di'sajn/ er det ikke godt å si om <g> inngår i digraf med <ig>=/aj/ eller <gn>=/n/ eller er "stum" <g>=/Ø/. I slike ord har bokmål importert hele det grafonomiske tegnet fra engelsk, og kanskje kan det best analyseres som en trigraf <ign>=/ajn/ som i engelsk.

I adjektiv som *vennlig*, veksler <g> mellom null og /k/, mens i nøytrums-substantiv veksler <t> mellom null og /t/. Også <v> og <d> veksler mellom null og herdet konsonant, men det er vanskelig å finne veksling mellom null og ustemt konsonant i samme rot:

<vennlig>=/²venli/ ~ <vennligst>=/¹venlikst/

<huset>=/¹hu:se/ ~ <husets>=/¹hu:sets/

<gav>=/ga:/ ~ <gave>=/²ga:ve/, <gjev>=/je:v/ ~ <gjevt>=/je:ft/

<bad>=/ba:/ ~ <bedende>=/²bedne/ ~ <bedt>=/bet/

<Gud>=/gu:d/ ~ <Guds>=/guts/

(Null-lisensiering av finnes ikke i bokmål, slik det gjør i engelsk <lam>=[læm], <bomb>=[båm]. De historisk eliderte /b/'ene blir ikke representert i bokmål, slik eliderte /d/'er, /g/'er og /v/'er blir.) Det som imidlertid er mest påfallende i disse vekslingene, er vekslingen mellom null og ustemt konsonant for <t>. Ingen andre grafemer med ustemte konsonanter som primærfonem (<p>, <k>, <f>) har slik veksling, så det hadde vært mer i tråd med resten av systemet å la <d> representere vekslingen i nøytrum:

/¹hu:se/=*<hused> ~ /¹hu:sets/=*<huseds>

Dette hadde også ført til at en del ikke-homofone homografer hadde blitt oppløst:

<funnet>=/²funnt/ : *<funned>=/¹fune/

<vannet>=/²vannt/ : *<vanned>=/¹vane/

5.4 Konsonantsekvenser og digrafer

<i>graf</i>	<i>fonem</i>	<i>fremmed</i>	<i>eksempler lange</i>	<i>eksempler korte</i>
skj	ʃ		skjul	
sj	ʃ		sjø	æsj hasj
kj	ç		kjøl	bikkje
tj	ç tj		tjern tjene	
ng	ŋ ŋ			tang finger hangar tango
nk	ŋk			tank lenke
gn	ɲn _# nj gn		sagn Dagny vignett Dagny gnu	gnu Dagny
sk _(i,y,ei,øy)	ʃ		ski sky skei skøyte	
rs _#	ʃ			
rt	t		sart	fort
rd	d t r rd		?burde bord	?burde burte verden lærd
rn	ɲ		ørn kjerne	
rdt				hardt
kkje				bikkje

Kontekstavhengige representasjoner kan også tolkes som polygrafer med fonemsekvenser som innhold. De fleste av disse har også representasjoner som følger linearitetsprinsippet.

I enkelte tilfeller kommer digrafer i konflikt, f.eks. i <negl>=n+eg+l eller n+e+gl. Dette er imidlertid sjelden, hovedsakelig fordi de færreste digrafer involverer vokaler. De fleste slike konflikter finner derfor sted over junktur, som i *buskjul* og *brekkjern*.

I tilfeller som <vondt>= v+o+nd+t eller v+o+n+dt spiller det ikke noen rolle, for primærfonemene og linearitetsprinsippet gir det samme resultatet uansett: /vont/. Man kan imidlertid tenke seg at det finnes et hierarki av digrafer, der enkelte digrafer har høyere prioritet enn andre, slik som addisjon og multiplikasjon i matematisk notasjon (4.1.2). I <England> finnes en potensiell digrafkonflikt mellom <ng> og <gl>, en konflikt som kanskje kunne løses ved et prioritethierarki. Fonotaktiske begrensninger gjør imidlertid at bare den ene lesningen er mulig, så vi får ikke teste hypotesen her.

I <egne>=¹e:gne/, ²æjne/, ²eɲne/ er det flere digraftolkninger. Her er alle lesningene gyldige, så her er det ingen konflikt og ingen prioritet. I <tegn>=/tæjn/, derimot, genererer de andre digrafene ikke gyldige ordformer. Dette tyder på at det kun er leksikalske egenskaper ved ordformene som er bestemmende for hvilke digrafer som skal ha prioritet, og ikke noen iboende egenskap ved digrafene.

<i>graf</i>	<i>fonem</i>	<i>fremmed</i>	<i>eksempler</i>	<i>eksempler fonemsekvens</i>
ski	ʃi		ski	BMOB ¹
sky	ʃy		sky	BMOB
skei	ʃæj		skeis skeie	BMOB
skøy	ʃøj		skøyte	skøyer
ki	çi ç		kiste kiosk	kivi kis
ky	çy		kyse	kybernetikk kykeliky
kei	çæj		keitet	keiser
køy	çøj		BMOB	køye
gi	ji		gi	gir
gy	jy		gyse	gylf
gei	jæj		geit	Geir geigerteller
gøy	jøj		gøys	gøy

¹Ingen moteksempler i BMOB.

Et spesialtilfelle av denne typen er der det ene grafemet i sekvensen har nulluttrykk i denne posisjonen. Også disse tilfellene kan analyseres på to måter, enten som digrafer eller som enkeltgrafem med nulluttrykk i visse posisjoner. Uansett hvilken analyse man velger, er det slik at grafemsekvensens primære fonologiske uttrykk i stor grad er en funksjon av fonotaktiske regler i SØ.

<i>grafem</i>	<i>fonem</i>	<i>fremmed</i>	<i>digraf-eksempler</i>	<i>lineære eks.</i>
hv	v		hvit	
hj	j		hjul	
gj #_	j		gjel	BMOB
lj #_	j		ljå	delje detalj
gl _#	l		fugl	negl hagl
nd	n		rund runde,v	runde,s
ld	l		ild	Hilde
rl	l		jarl erle	

Det er ikke vanlig å tolke <r>|_l=/Ø/, men dette er en konsekvens av analysen av retrofleks lateral som en allofon av /l/.

Også en del grafemsekvenser med primært fonologisk uttrykk som er fonemsekvenser, men som avviker fra linearitetsprinsippet, skriver seg fra fonotaktiske regler i SØ.

<i>grafem</i>	<i>fonem</i>	<i>eksempler</i>
sl	ʃl	slå vesle
nk	ŋk	tank renke
gn	ŋn	sagn

Disse kan kanskje best analyseres som enkeltgrafemer, der fonemsekvensen som genereres av linearitetsprinsippet, deretter omformes av det fonologiske regelsystemet:

{sl} → /ʃl/

{nk} → /ŋk/

$$\{\text{gn}\} \mid \neg\#_ \rightarrow / \eta \text{n} / ^1$$

Dette er regler som virker betingelsesløst med unntak av over morfemgrense:

$$\{\text{russ\#land}\} \rightarrow / ^1 \text{ruslan} / , ? / ^1 \text{ru}\text{ʃ} \text{lan} /$$

$$\{\text{vann\#kur}\} \rightarrow / ^2 \text{vanku:r} /$$

$$\{\text{leg\#ning}\} \rightarrow / ^2 \text{legni}\eta /$$

5.5 Allofoni

Som nevnt over, er ikke markering av allofoni noe systematisk trekk ved bokmål. Punktene under forutsetter en tolkning av SØ som gjør at de variasjonene som diskuteres, faktisk regnes som allofoner.

<y> i diftong er en markør for rounding av sluttpunktet i diftongen, noe vi har analysert som en allofonisk variasjon. Markøren er relativt sterk og nesten total, men viser svakhet i <bayer>=¹bajer/ og partialitet i <koie>=²kåje/, <jojo>=²jojo/ og <huie>=²huje/, som alle trolig har allofonisk rounding av diftongens sluttpunkt.

<r>|_1 er en sterk men partiell markør for retrofleks lateral. Ifølge vår analyse av [l] som en allofon av /l/ er dette da en allofonisk markør.

5.6 Prosodemer

Prosodemet trykk blir ikke omtalt her.

5.6.1 Lengde

Kvantitetsfenomenet i SØ er relativt komplisert. Som nevnt over er en stavelse enten lang eller kort. Uaksentuerte stavelser er korte, mens aksentuerte stavelser er lange eller overlange.

¹Kontekstregelen i dette uttrykket sikrer riktig uttale i ord som <gnu>=/gnu:/.

Når vi snakker om plassering av trykk i en ordform eller et leksem i SØ, mener vi egentlig plassering av det realiserte trykket ved en eventuell aksentuering av ordet i en ytring. Uaksentuerte ord har ikke noe fonologisk trykk. Imidlertid endrer ikke rettskrivningen av en ordform seg med hensyn til om ordet er aksentuert eller ikke¹, så når vi studerer kvantitetsmarkeringen i ordformer er det den *potensielle* trykkplasseringen vi går ut fra.

Lange stavelser har enten lang vokal eller lang utlydskonsonant. Lang utlydskonsonant kan bestå i en geminert konsonant, som i [hat:]=<hatt>, eller en konsonantsekvens, som i /hast/=<hast>. Overlange stavelser består av lang vokal og i tillegg lang utlydskonsonant. Dette er vanligvis bøyde former, som i [bi:lt] =<bilt>="av *bile*, v.", men det finnes også enkelte ubøyde former med overlange stavelser, som i <påske>=[²på:skə].

I noen tilfeller av herding av stemte konsonanter vil lisensiering av grafem for stemt fonem markere for vokalkvalitet. Det finnes minimale par som:

<stivt>=/sti:ft/ : <stift>=/stift/
<lovt>=/lå:ft/ : <loft>=/låft/
<tregt>=/tre:kt/ : <trekt>=/trekt/²

Det finnes imidlertid mange moteksempler her, slik at dette er ikke et regelmessig trekk:

<sagt>=/sakt/ : <sakte>=²sakte/
<bredt>=/bret/ : <brett>=/bret/
<lyst>=/ly:st/ : <lyst>=/lyst>

¹ Et mulig unntak er ordformene <én>, <éi>, <ett>. Disse formene kan tolkes som at de signaliserer aksentuert lesing av artiklene <en>, <ei>, <et>, men en vanligere tolkning er at <én>, <éi>, <ett> representerer *tallordene* /e:n/, /ei/, /et/, mens <en>, <ei>, <et> representerer *artiklene*, altså slik at aksentene skiller mellom ulike leksemer i ulike ordklasser. Se kapittel 5.8.3.

²<trekt> er ikke med i den kodifiserte bokmålsnormen, men likevel ganske vanlig i usus. Usenet-korpuset har omtrent 10% forekomster av <trekt> i forhold til <trukket>, som er i den kodifiserte normen (256 mot 2320). I tillegg kommer former som <trekkt> (7), <trukki> (4) og <trekki> (2).

Se ellers omtalen av digrafer og funksjonelle enheter i kapittel 4.5.8.

5.6.2 Tonem

Distinktiv tonemforskjell i SØ kan bare realiseres i sammenhenger med to påfølgende stavelser, som i /²bøner/= <bønner> : /¹bøner/= <bønder>, men disse to stavelsene kan godt være i to ulike påfølgende ord, som i /¹ta:på/= <ta på>="berøre" : /²ta:på/= <ta på>="kle på seg".

Det finnes ikke noe system i rettskrivningen som skiller mellom tonem verken innen ordformer eller over ordgrenser. Derimot finnes det mange enkeltstående stavemåter som skiller mellom minimale (og subminimale) par der tonem er distinktivt:

1. <vanne>=/¹vane/ : <vanne>=/²vane/
2. <bønder>=/¹bøner/ : <bønner>=/²bønner/

Imidlertid er det her fullt av fallgruver; i 1) over kan vi også ha <vanne>=/²vanet/, mens det minimale paret <baner>=/¹baner/ : <baner>=/²baner/ er homografer.

5.7 Stavelsen

Stavelsen er et intuitivt lettfattelig men teoretisk komplisert begrep. Vi skal ikke gå inn i en diskusjon omkring stavelsens fonologiske eller fonetiske natur her, men følger Vanvik 1979.

5.7.1 Antall stavelser

I SØ kan vi telle stavelser i ord ved å telle auditive lydstryketopper for stemte lyder. Stavelseskjernen er oftest en vokal, men kan også være følgende nasaler og likvider: /n ŋ l ʎ r/. Eksempler er

¹Vanviks fremstilling analyserer retrofleks lateral som eget fonem, og her følger vi den analysen. Det er ingen konsekvenser i denne sammenhengen.

- <banan> /ba¹na:n/
- <båten> /¹bå:tɲ/
- <arten> /¹a:tɲ/
- <seidel> /¹sæjdl/
- <størrelse> /²stø|se/
- <lærer> /²læ:r/

I skrift er stavelseskjernen alltid markert ved et vokalgrafem. Det betyr at om vi teller antall vokalgrafemer i en ordform, finner vi en sikker øvre grenseverdi for antall stavelser i ordformen. Der stavelseskjernen ikke er realisert gjennom en vokal, vil vokalgrafemet som representerer stavelseskjernen normalt være <e>, ettersom det bare er schwa som elideres på denne måten:

<råtne>=²råtne/ : <rottene>=²råtne/

Enkelte fremmedord vil imidlertid kunne ha andre vokaler:

<badminton>=¹bædmintɲ/, /¹bædmintɲ/

<remington>=¹remintɲ/, ?/¹remintån/

<tartan>=?/¹ta:tɲ/="skotskrutet stoff"¹

I noen tilfeller av fremmedord kan også andre sekvenser representere stavelsesbærende konsonant:

<double>=¹døbel/, /¹døb|/

Her har vi

<e>={stavelse}|l_#

Dette er en representasjon som er importert fra henholdsvis engelsk og fransk sammen med ord av denne typen. I alle tilfellene er det likevel elisjon av schwa som ligger til grunn.

¹Tartan kan nok oppfattes som et kodeskifte i norsk.

Imidlertid kan antall stavelser gjerne være lavere enn antall vokalgrafemer. Dette kan ha flere ulike forklaringer:

1. Ordet kan ha gjennomgått en diakron elisjon. I slike tilfeller har ordet *alltid* den eliderte uttalen, med mindre det kan få en leseuttale med høyere antall stavelser. Et eksempel er <elleve> = /²elve/.
2. Ordet kan elideres synkront. I slike tilfeller finnes gjerne i tillegg en variant i en høy stil med det antall stavelser som stemmer overens med antall vokalgrafemer. Et eksempel er <værelse> = /²væ:lse/ med høystilsvarianten /²væ:relse/ eller /²væ:r|se/.
3. To vokalgrafemer kan representere en diftong. Dette gjelder bare grafemsekvensene <au>, <ai>, <ei>, <øy>, <oi>, <ui>, men disse sekvensene kan også representere monoftongsekvenser. Dette forekommer helst i morfologisk komplekse ordformer, som avledningen <reing> eller sammensetningen <snøyr>. Disse tilfellene kan ikke skilles grafonomisk.¹

Heller ikke dette prinsippet gjelder uten unntak. Eksempler på ord med flere stavelser i det fonologiske uttrykket enn vokaler i det grafonomiske uttrykket, er <hagl>=/¹hagəl/, <kakl>=/¹kakəl/, <jugl>=/¹jugəl/, <tagl>=/¹tagəl/, <putl>=/¹putəl/, foruten imperativformer som <formidl>=/får¹midəl/, <sykl>=/²sykəl/, /²syklə/.

I alle disse eksemplene gjør formen som ikke bruker en vokal til å representere den siste stavelsen, at prinsippet om morfemets likhet følges: I imperativformene brukes stammeformene, som også brukes i infinitiv (og for svake verb også for andre bøyingsformer), mens for de fleste andre eksemplene representeres morfemet på samme måte som i verbformene <hagle>, <kakle>, <jugle> og <putle>, enten disse er avledet av substantivene eller motsatt. For <tagl> finnes det imidlertid ikke noen avledet form, så her finnes det ikke noe slikt argument for ikke å stave ordet *<taggel>=/¹tagəl/, parallelt med for eksempel

¹ Enkelte ganger brukes <-> i sammensetninger der første ledd har vokal i utlyd og andre ledd har vokal i fremlyd, for å lette lesingen, som for eksempel i <bre-is>.

<nøkkel>=/²nøkel/. Også for <jugl>, vil verbformen <jugle> spille svært liten rolle, ettersom den er svært perifer og ikke forekommer i Oslo-korpuset i det hele tatt. I Oslo-korpuset finner vi da også bare én forekomst av <jugl>, men fem av <juggel>. I Usenet-korpuset er tendensen enda tydeligere; der finner vi ingen forekomster av <jugl>¹ og 74 av <juggel>.

<gl># er interessant i bokmål fordi sekvensen kan representere fire ulike fonologiske mønstre:

1. <gl>=/gel/ som i <hagl>=/¹hagel/
2. e<gl>= æ/jl/ som i <negl>=/næjl/
3. <gl>=/l/ som i <fugl>=/fu:l/
4. <gl>V=/gl/ som i <megle>=/²megle/

I <egl> er det digrafen <eg> som kan få prioritet over konsonantsekvensen <gl>, som bare har en svak karakter av digraf, men vi ser av punkt 4 at denne prioriteten ikke er entydig. Det er også interessant å merke seg at sekvensen <ggel> ikke finnes i *Bokmålsordboka*, slik at det kan sees på som regelmessig å notere /gəl/ uten stavelsesrepresenterende vokal: <gl>.

5.7.2 Stavelsesgrense

Stavelsesgrense kan være betydningsskillende i SØ. Vi har for eksempel et minimalt par som *u-tøy* : *ut-øy*. Fonetisk skilles disse ordene ved aspirasjon av plosiven.

5.8 Intonasjon

Setningsintonasjon i vid forstand reflekterer talerens holdning til verden, talesituasjonen eller ytringen. Den er normalt ikke betydningsskillende på ordnivå, som er det høyeste nivået vi behandler i denne oppgaven. Det er imidlertid enkelte unntak. Problemstillingene på nivå over ordnivå ligger

¹ Bortsett fra to treff i et innlegg om genealogi, der <Jugl> ser ut til å referere til et geografisk navn.

egentlig utenfor emnet for denne oppgaven; dette gjelder generelt, men spesielt tegnsetting.

5.8.1 Tonemgruppe

En tonemgruppe kan gå over to ord, og slik sett kan det betydningsskillende trekket tonem angå enheter på et høyere nivå enn ordet. Eksempler er :

<ta på>frakken=¹ta:på/="berøre" : ²ta:på/="kle på seg"

Slike minimale par er homografer.

5.8.2 Aksentuering

Aksentuering av ulike ord i en setning kan være betydningsskillende på to måter.

Ulik inndeling i intonasjonsgrupper kan gi ulik setningsbetydning uten at hvert enkelt ords betydning endres:

- *Han lovte-ikke å-komme : Han-lovte ikke-å-komme*

Plassering av trykk kan også endre et ords betydning i sammenhengen, eventuelt endre referansen til et annet ord i et homofonpar .

- *Han ¹tok nok (= "trolig") et ¹glass : Han tok ¹nok (= "enda") et ¹glass*
- *Han ¹drakk nok (= "trolig") ¹øl : Han drakk ¹nok (= "tilstrekkelig") ¹øl*

Aksentuering markeres normalt ikke i skrift, men det finnes enkelte unntak, der aksentuerte ord, særlig <det>, kan markeres med en akutt aksent for å lette lesingen. Dette skal normalt gjøres kun der lesingen blir betydelig vanskeliggjort om aksenten ikke brukes, ifølge Vinje 2004.

5.8.3 Tallord og artikler

Det sies gjerne at akutt aksent skiller mellom de ubestemte artiklene eller pronomenene på den ene siden og de minste tallordene på den andre:

<en> : <én>

<ei> : <éi>

<et> : <ett>

I nøytrumsparet er skillet realisert gjennom enkel eller dobbel konsonant.

Det er imidlertid ikke slik at tallordene *må* markeres på denne måten, og vi kan da stille spørsmålet om aksenten, henholdsvis dobbeltkonsonanten, egentlig representerer aksentuering i ytringen.¹ For maskulinum- og femininumformene er dette i så fall en parallell til markeringen av ord som <dét> med et diakritikum ved aksentuering, men for nøytrumsformen vil dette være et enestående tilfelle i rettskrivningen der et ords stavemåte endres med aksentueringen av det i en ytring.

5.8.4 Pauser

Heng ham, ikke vent til jeg kommer : Heng ham ikke, vent til jeg kommer

Dette kjente eksemplet viser hvordan komma kan være betydningsskillende på setningsnivå. Kommaet kan her sies å representere grensen mellom to grammatiske enheter, nærmere bestemt setninger, men ettersom det er vanskelig å gå rett fra det fonetiske uttrykket til setningsanalyse, kan kommaet i hvert fall i dette tilfellet også sies å representere den pausen som normalt vil forekomme i disse ytringene.

5.8.5 Spørsmål

Du kommer. : Du kommer?

¹Faarlund, Lie og Vannebo 1997 bruker ikke skolegrammatikkens ti ordklasser, men klassifiserer både tallordene og de ubestemte artiklene som determinativer av typen kvantorer. De skiller likevel mellom tallordet *én* og artikkelen *en* (s.217) og sier (s.224) at ordet "gjærne" er trykksterkt i den første betydningen og trykksvak i den siste. Men deretter sier de: "I skrift blir det av og til brukt aksent over e-en i maskulinum og femininum på bokmål for å markere at ordet er trykksterkt." Dette holder muligheten åpen for begge analysene. Vinje 2004 sier at "Aksenttegnet tydeliggjør skillet mellom tallordet *én* og den ubestemte artikkelen *en*."

Her skiller bruk av punktum henholdsvis spørsmålstegn mellom en utsagnssetning (eller kommando) og et spørsmål. Dette skillet vil gjerne være realisert gjennom stigende setningsintonasjon i spørsmålet og synkende i utsagnet, men dette er en langt mer komplisert materie enn at vi kan gå inn i den her. Eksemplet viser bare at tegnsetting kan skille mellom ulike setningsbetydninger i setningspar der alle ord og dessuten ordrekkefølgen er identiske.

6 Andre perspektiver på bokmål

6.1 Syllabografi

Vi skal ikke for alvor påstå at bokmål er et syllabografisk skriftsystem, men vi skal vise at enkelte prinsipper i bokmål har paralleller i stavelsesbaserte systemer, i hvert fall om vi studerer sammenhengene i leseretning.

Mange viktige skriftsystemer, både historiske og nåtidige, er syllabografiske. I slike systemer representerer prinsipielt ett grafem én stavelse, slik tilfellet er i japansk kana. Et språk med komplisert stavelsesstruktur vil vanskelig kunne representeres godt ved et rent stavelsesbasert system, rett og slett fordi det ville kreve for mange grafemer, og det finnes flere typer tilpasninger som kan være nødvendige for å oppnå tilstrekkelig økonomi i systemet avhengig av språkets stavelsesstruktur. Ulike typer tilpasninger av prinsippet finnes. I eksemplene under er stavelsesgrafemer representert ved sekvenser av de konsonanter og vokaler de representerer.

$$[KVK] = \langle KV_1 + V_1K \rangle$$

$$[KVK] = \langle KV_1 + KV_1 \rangle (+\text{nøytraliseringsgrafem})$$

$$[KVK] = \langle KV_1 + K \rangle$$

$$[KVK] = \langle KV_1 + KV_{\text{nøytral}} \rangle$$

$$[KKKVK] = \langle KV_1 + KV_1 + KV_1 + K(V_1) \rangle$$

Ulike varianter eller kombinasjoner av disse typene tilpasninger finnes også.¹

6.1.1 Fremlyd og utlyd

Vi har i kapitlet om bokmål som fonemografisk skrift, sett på de ulike mulige grafemiske representasjonene av de ulike fonemene i SØ, og vi har sett at det er begrensninger med hensyn til i hvilke omgivelser de ulike representasjonene kan

¹Sampson 1985.

brukes; for eksempel <hv>|#_=/v/, mens <rs|_#=/ʃ/. Begrensningene har ofte etymologisk bakgrunn eller begrunnelse, men med en synkron tolkning av skriftsystemet, finner vi mønstre som kan tolkes med stavelsen som ramme. Det er særlig to problemer vi må være oppmerksomme på om vi skal bruke stavelsen som en enhet i en analyse av skriftsystemet:

1. Det er i SØ ikke entydige stavelsesgrenser, slik det er i mange av de språkene som bruker stavelsesbaserte systemer, f.eks. japansk.
2. Stavelsesgrensene faller i SØ ofte, men ikke alltid, sammen med morfemgrensene.

Vi må selvfølgelig også ha de fonotaktiske restriksjoner på stavelsesstrukturen i SØ i mente; <ng> kan ikke representere /ŋ/ stavelsesfrontalt fordi dette fonemet ikke kan opptre stavelsesfrontalt. Men om vi ser på fonemer med mer komplett distribusjon, som /n/ og /ʃ/, finner vi grafotaktiske begrensninger som har stavelsen som ramme. Følgende hypotetiske stavelser som følger de fonotaktiske reglene i SØ tjener som eksempler:

- /nen/ = <nenn>, <nend>, */nnen/, */nden/
- /ʃa:ʃ/ = <skjars>, <sjars>, */rsars/, */skjaskj/, ?/skjasj/

<sj># signaliserer antagelig kort vokal, som i <hasj>, <isj>, <krasj>. Det finnes imidlertid flere importord, særlig fra fransk, med trykktung sistestavelse med lang vokal og <sj>, f.eks. <fromasj>, <kollasj>. I bokmålsordboka er faktisk alle ord med <sj># importord, bortsett fra interjeksjoner som <æsj> og <isj>, pluss <hesj>¹ (av <hesje>) og <bæsj>, alle med korte vokaler. Også enstavelses importord med <sj># har kort vokal; det gjelder <tusj>, <dusj> og <pysj>. Ingen av disse kan sies å være fremmedord.

¹ NUOB og NUOV har her /heʃ/, mens min intuisjon sier /hæʃ/.

6.1.2 Markering av vokalkvantitet

Den vanlige tolkningen av kvantitet i SØ er knyttet til stavelsen, som f.eks. i Sivertsen 1967 og Haslev 1985. Ifølge denne tolkningen er aksentuerte stavelser lange, og lengden er plassert enten i vokalen eller i den etterfølgende konsonanten: V:K : VK: . I en fonemisk analyse regnes imidlertid bare vokalkvantiteten som distinktiv, mens konsonantkvantiteten bestemmes av omgivelsene.¹ Et minimalt par er:

/vi:n/=<vin> : /vin/=<vinn>, <vind>

Her er det den doble konsonanten <KK>, i disse tilfellene den ekstra <n>, henholdsvis <d>=/Ø/, som markerer kort foregående vokal. Som vi ser i Venezky 1970, kan dette tolkes som at den ekstra konsonanten her opptrer som en markør. I et syllabisk perspektiv signaliserer den ekstra konsonanten en modifikasjon av stavelseskjernen som (det syllabiske) systemet ikke er i stand til å signalisere på egen hånd, nemlig forkorting av vokalen. Det vil si at det umarkerte er lang vokal i trykksterk stavelseskjerne.

6.1.3 Markering av antall stavelser

Et annet fenomen som også er behandlet ellers i oppgaven², kan også sees i lys av et syllabisk prinsipp. Alle stavelser må markeres utvetydig med et vokalgrafem i stavelseskjernen, uavhengig av om stavelseskjernen er realisert som vokalfonem eller ikke:

/²rātne/=<rottene> : /²rātne/=<rotne>

¹ Ifølge Vanvik 1979 er "lang" konsonant heller ikke entydig lengre rent fonetisk, og han hevder på bakgrunn av dette at det er vokalens kvantitet som uansett må regnes som det distinktive trekket.

² Se 3.1.1.

Denne representasjonen er imidlertid entydig bare i én retning, nemlig fra fonologisk uttrykk til grafonomisk uttrykk. I motsatt retning vil det kunne finnes vokalgrafemer som aldri vil realiseres som stavelseskjerner, f.eks.

- <elleve>=/²elve/, */²eleve/
- <værelse>=/²væ:lse/, ?/²væ:relse/

I disse tilfellene må vi søke forklaringsmodeller andre steder enn i stavelsen.

6.1.4 Markering av stavelsesgrensen

Bokmålsnormen åpner for ved behov å markere stavelsesgrense med bindestrek for å lette lesingen, men bare der stavelsesgrensen faller sammen med morfemgrense, i praksis ved sammensetning. At bindestreken løser opp et homografipar er ikke nødvendig; det kan dreie seg om ord som er vanskelige for leseren å parse om man ikke finner junkturplasseringen og altså stavelsesgrensen, for eksempel fordi feilplassert junktur trigger tolkning av en grafemsekvens som en digraf:

- <bre-is> blokkerer lesing med diftong: <breis>=*/bræjs/. I dette tilfellet markerer bindestreken også *antall* stavelser, og ikke bare grensen.
- <trafikk-kork> blokkerer at begge <k>-ene tilegnes førsteleddet, som jo skal ha to <k>-er, slik at det ikke blir noen igjen til sisteleddet:
<trafikkork>=*/tra¹fik-årk/

6.2 Morfemografi

Mange fonemografiske skript generelt og bokmål spesielt forskjellige egenskaper som har karakter av morfemografi:

1. Automatiske variasjoner i fonologisk kontekst, som i engelsk "flertalls-s", der /s/ representerer flertallssibilant uavhengig av hvordan denne realiseres: <pats>=/pæts/, <pads>=/pædz/. I bokmål finnes eksempler som <stiv>=/sti:v/, <stivt>=/sti:ft/.

2. Leksikalske markører. Det finnes tre varianter av leksikalsk markering: som gjør homofone eller segmentalhomofone par grafemisk ulike: eller
 - a. Fullstendig homofoni, f.eks.: <hva>=/va:/ : <vad>=/va:/
 - b. Segmentalhomofoni, f.eks. <vannet>=/¹vane/ : <vanne>=/²vane/.
 - c. Ikke-homofoner der f.eks. digrafskopus eller andre regler er i konflikt: <kikk>=/çik/ : <kick>=/kik/
3. Morfemets konstans. Morfemets stavemåte opprettholdes uavhengig av uttaleendringer i ulike omgivelser gjennom bøyning, avledning eller sammensetning. Det finnes to varianter:
 - a. Identisk staving:
 - i. <seks>=/seks/ ~ <seksten>=/²sæjstn/
 - ii. <god>=/go:/ ~ <godt>=/gåt/
 - iii. <der>=/dæ:r/ ~ <dersom>=/¹dæfåm/
 - b. Ikke-identisk staving, men med en markør i vokalgrafemet:
 - i. <så>=/så:/ ~ <sæd>=/se:d/, ikke *<sed>, som ville ha vært den fonemografiske representasjonen, og ikke *<såd>, som ville ha vært den morfemografiske representasjonen;
 - ii. <syv>=/sy:v/ ~ <sytten>=/²søtn/, ikke *<søtten>, som ville ha vært den fonemografiske representasjonen, og ikke *<syvten>, som ville ha vært den morfemografiske representasjonen.
4. Konvensjonell representasjon av bøyings- og avledningsmorfemer, der representasjonen i liten grad er basert på uttale, særlig nullmorfemer i bokmål:

- a. <eplet>=²eple/, der <t> representerer bøyningsmorfemet.
- b. -lig=<lig>=/li/ der <g> markerer adjektivavledningen.

Disse representasjonene kan imidlertid også analyseres som automatiske variasjoner (variant 1 over) eller morfemografisk konstans (variant 3 over) i og med at vi har vekslinger mellom nulluttale og plosiv:

- a. <eplet>=²eple/ ~ <eplets>=²eplets/
- b. <vennlig>=²venli/ ~ <vennligst>=¹venlikst/

Suppletive bøyningsformer blir alltid representert fonemografisk:

<god>=/go:/ ~ <bedre>=¹be:dre/ ~ <best>=/best/

6.2.1 Vokalveksling

Vokalveksling som bøyings- eller avledningsmekanisme eller som automatiske vekslinger er relativt utbredt i norsk, skjønt kanskje ikke så utbredt som i en del andre språk, f.eks. engelsk [ai] ~ [i] i *sign* ~ *signature* og [ei] ~ [ə] i *information* ~ *informative* eller arabisk *kitab*="bok", *kutub*="bøker", *katab*="han skrev" eller russisk <города>=¹gorədΛ/"en bys", /gərΛ¹da/"byer".¹

I bokmål representeres som regel vokalvekslingen fonemografisk, som i

/²syŋe/=<syng> ~ /saŋ/=<sang> ~ /²soŋet/=<sunget>.

I vekslingen /å/ ~ /e/, /æ/ er det imidlertid ikke slik. I ord med /e/ eller /æ/ som er avledet² av /å/, lisensierer vokalen <æ>, uavhengig av hvilken vokal det er:

- /tå:/=<tå> ~ /tæ:r/=<tær>
- /få:/=<få> ~ /¹fære/=<færre>³
- /så:/=<så> ~ /se:d/=<sæd>

¹Engelske eksempler fra Jones 1991, arabiske fra Haslev 1972, russiske fra Sproat 2000. De russiske eksemplet er omtalt i mer detalj i 4.5.5 og 4.5.6.

² Avledet er brukt i utvidet betydning her og står for både bøyning og avledning.

Dette er imidlertid i utpreget grad et diakront prinsipp. Ordet med <å> vil ofte være ukjent for språkbrukeren, eller i det minste er den historiske forbindelsen ukjent eller ubevisst. For eksempel tror jeg et mindretall av språkbrukere uten filologisk utdannelse kjenner de etymologiske forbindelsene under.

- <skjæker>=²ʃe:ker/ ~ <skåk>=/skå:k/
- <færing>=²fæ:riŋ/ fra <åre>=²å:re/

Noen enkeltstående tallord har vokalveksling i stammen og stavemåte som avviker ganske markert fra det fonemografiske prinsipp:

- <seksten>=²sæjstn/, altså med korrespondansen <ek>=/æj/.
- <syttten>=²søtn/ og <sytti>=¹søti/, altså med korrespondansen <y>=/ø/.
- <tredve>³=²trædve/⁴, altså med korrespondansen <e>=/æ/.

³ Dette er et kjent problemområde for skriftspråkbrukere, også voksne. Dessuten finner vi ofte <verre> feilstavet som <værre>, noe som ganske sikkert er etter analogi med <færre>. Usenet-korpuset har 2875 forekomster av <værre> mot 11933 av <verre>. Dette er en feilprosent på 19, som er den høyeste feilprosenten jeg har funnet i arbeidet med denne oppgaven.

³<tredve> er ikke med i den kodifiserte bokmålsnormen, men er klart med i den operative normen. Oslo-korpuset gir 112 treff på <tredve> og 99 på <tretti>.

⁴ NUOB har både /æ/ og /e/ her, mens NROB og Vanvik (1979) har /æ/.

<tredeve>=^{/2}trædve/ og <tredevte>=^{/2}trædefte/ er kanskje de eneste hjemlige ord med /æ/ foran noe annet enn /r/ eller retroflekser i SØ.¹ Det er derfor ikke mulig å si noe om hva som er den regelmessige representasjonen av /æ/ i denne posisjonen, men primærgrafemet til /æ/ er <æ> og ikke <e>. Lisensieringen av <e>=/æ/ kan derfor analyseres som en opprettholdelse av rotmorfemet <tre>=/tre:/. I <seksten> blir også rotmorfemet, <seks>, opprettholdt ved den svært uregelmessige korrespondansen <ek>=/æj/, mens i <sytten> og <sytti> er det bare vokalgrafemet som signaliserer slektskapet med <syv>. *<syvten> og *<syvti> ville ha opprettholdt rotmorfemet fullstendig, og det ville heller ikke innført noen ny nullrepresentasjon for <v>, ettersom denne allerede eksisterer i ord som <tolv>=/tål/ og til og med postvokalisk i <gav>=/ga:/ og svak form av <av>=/a(:)/, dog ikke i utlyd postkonsonantisk.

Dette er imidlertid bare noen enkeltstående tilfeller og ikke noe prinsipp for tallordene; ^{/2}aten/ skrives <atten> og ikke <åtten>, som man kanskje kunne tenkt seg, og /førti/ skrives ikke lenger <firti> som i 1907-rettskrivningen (Aars 1907).²

6.2.2 Kvantitet og junktur

Ved junkturer oppstår gjerne grafemsekvenser som faller utenfor grafotaktiske regler. For eksempel gir <trafikk>+<kork> en sekvens på tre like konsonanter, og en konsonant må strykes eller en bindestrek settes inn for å oppnå gyldig grafotaks.

Sekvenser av typen <K₁K₁K₂> er også belastet i en grafotaktisk sammenheng. For eksempel gir nøytrumsbøyning av /tøf>=<tøff> ikke /tøft>=*<tøfft>, men <tøft>. Hovedregelen er at sekvensen forenkles til <K₁K₂> ved bøyningsjunktur, som i <vilt>=/vilt> og <spilt>=/spilt/. Imidlertid blokkeres regelen ved forekomst av et segmentalhomofon i leksikon, noe som gir

¹I tillegg blir en del tabuord gjerne uttalt med /æ/. En slik /æ/-uttale regnes normalt ikke som del av SØ, så det er trolig slik at /æ/ i disse ordene nettopp signaliserer lavstil.

²<firti> og <fyrti> var valgfritt.

$full+t \rightarrow /fult/= <fullt> \mid \leftarrow /fu:lt/= <fult> \in A_{\text{Leksikon}}$

$viss+t \rightarrow /vist/= <visst> \mid \leftarrow /vi:st/= <vist> \in A_{\text{Leksikon}}$

Denne regelen har den ulempen at den må overvåke leksikon for nyord for å virke. Den har også unntak, for eksempel

$spill+t \rightarrow /spilt/= <spilt> : spil+t \rightarrow /spi:lt/= <spilt>.$

I sammensetninger forenkles grafemsekvensen ikke, mens ved avledninger er hvert avledningsmorfem merket for $\{\pm\text{konsonantforenkling}\}$.

-ning <forflytte> ~ <forflytning>

-ke <tørr> ~ <tørke>

-ne <mugg> ~ <mugne>

-ske <grønn> ~ <grønske>

-lig <venn> ~ <vennlig>

-het <glatt> ~ <glatthet>

-dom <mann> ~ <manndom>

I <forvissning> gjelder den leksikonbaserte unntaksregelen som for bøyde former, trigget av <forvisning>.

6.3 Et generativt perspektiv

6.3.1 Generativ grafonomi

Generativ fonologi er behandlet av Kristoffersen 2000, Katamba 1989, Durand 1990 og Chomsky og Halle 1968.

Vi er her ikke først og fremst interessert i den generative fonologien, men i å vise hvordan rettskrivningen kan tolkes ut fra en generativ tankegang ved å postulere at stavemåtene gjenspeiler dypfonlogiske former. De underliggende formene vi

opererer med, er bare ment å skulle illustrere prinsippet, og de er dermed ikke nødvendigvis *morfofonemiske* enheter i noe eksisterende generativt rammeverk.

hurtig

Fonemografisk ville disse ordene ha vært skrevet *<hurti>, *<hurtikst>. Som vi har sett, kan <g> her tolkes som en markør for adjektiv, og vi har også det morfemografiske prinsippet som gir <g> i morfemet uavhengig av fonologisk form, som veksler mellom <Ø> og <k>, og annen G/F-kontekst.

1. *hurtig* + # → *hurti* | {g} strykes i trykksvak utlyd
2. *hurtig* + *st* → *hurtigst*
3. *hurtigst* → /hurtikst/ | {g} herdes før ustemt kontinuant {s}
4. *hurtig* + *ere* → *hurtigere*
5. *hurtigere* → /hurtigere/ | ingen herdning eller strykning

Det grafonomiske uttrykket er generativt og gjengir de underliggende formene, men uttalen /hurtigere/ kan ikke genereres av disse reglene.

huset

Fonemografisk ville disse ordene ha vært skrevet *<huse>, <husets>, selv om formen *<huse> sterkt signaliserer en uttale med tonem 2 og også gjør ordformene *huse*, *huset* til homografer, selv om de bare er partielle homofoner. Som vi har sett, kan <t> her tolkes som en markør for bestemt form nøytrum entall, og vi har også det morfemografiske prinsippet som gir <t> i morfemet uavhengig av fonologisk form, som veksler mellom <Ø> og <t>, og uavhengig av omgivelsene som gir disse ulike formene. Vi kan også tolke <t> som en markør for tonem 1, men dette er ikke noe underliggende prinsipp i skriptet, jamfør *funnet*, *kunnet*, *eplet*; det er snarere en bieffekt av at <et>, <t> markerer bestemt form entall, skjønt i tilfeller som *eplet*, der grunnformen har tonem 2, spiller heller ikke det noen rolle.

1. *hus* + *et* → *huset*

2. *huset* → /huse/ | t i utlyd strykes dersom den er (del av) bestemt-form-nøytrum-entall-morfemet.
3. *huset* + *s* → *husets*
4. *husets* → /husets/ | ingen strykningsregel virker.

blåbær

Fonemografisk ville dette ordet ha vært skrevet <blåbbær> eller <blobbær> om vi antok at ordformen er synkront opak. Selv om dette nok er en tvilsom antagelse, og folk flest er bevisst sammenhengen mellom *blåbær* og *blå*, kan vi vise at den generative modellen holder også i slike tilfeller.

- *blå:* + *bær* → *blå:bær* | forkortning av lang vokal i trykksvak stavelse
- *blå:bær* → /blåbær/ | forkortning av lang vokal i trykksterk stavelse i sammensetninger som ikke har noe produktiv preg. Jfr /¹solbær/, /¹båtsman/¹

6.3.2 Generativ lesing

I forrige underkapittel brukte vi et generativt apparat med underliggende, kanskje hypotetiske, fonologiske former som vi antok kunne ligge til grunn for stavemåtene vi undersøkte. I dette kapitlet skal vi se på andre eksempler og prøve å sette opp et apparat for å generere riktig uttale ut fra stavemåten, og vi vil se at vi trenger mer informasjon enn stavemåten for i et hvert tilfelle å kunne gjøre dette entydig.

Først et eksempel på et ord som er fullstendig fonemografisk:

¹ Ingen av disse eksemplene er helt parallelle til tilfellet *blåbær*. Både *solbær* og særlig *båtsmann* har konsonantsekvenser som vil ha påvirket den forutgående vokal i kort retning, i hvert fall i tidligere stadier av norsk. Men *stolben*, som fonologisk er ganske likt *solbær*, har beholdt sin lange vokal, noe som tyder på at det produktive i sammensetningen spiller en rolle her. Sammensetning med møbel + *ben* er produktiv (*sofaben*, *bordben*, ?*TV-møbelben*), mens sammensetninger med *bær* i mindre grad er det, i hvert fall for å denotere en bærsort. Sammensetninger med beskrivende framledd er fortsatt produktivt, som for eksempel *møkkabær*, *drittbær*, *kakebær*.

$$\begin{aligned} \langle \text{bil} \rangle &\rightarrow \{\text{bil}\} \mid \langle \text{b} \rangle \rightarrow \{\text{b}\} + \langle \text{i} \rangle \rightarrow \{\text{i}\} + \langle \text{l} \rangle \rightarrow \{\text{l}\} \\ \{\text{bil}\} &\rightarrow /bi:l/ \mid \{\text{VK}\# \} \rightarrow [\text{V}:\text{K}] \end{aligned}$$

For digrafer er det nødvendig med regler som tar grafemsekvenser som input:

$$\begin{aligned} \langle \text{kjøtt} \rangle &\rightarrow \{\text{çøtt}\} \mid \langle \text{kj} \rangle \rightarrow \{\text{ç}\} + \langle \text{ø} \rangle \rightarrow \{\text{ø}\} + \langle \text{t} \rangle \rightarrow \{\text{t}\} + \langle \text{t} \rangle \rightarrow \{\text{t}\} \\ \{\text{çøtt}\} &\rightarrow /çøt/ \mid \{\text{VK}_1\text{K}_1\}\# \rightarrow [\text{V}:\text{K}] \end{aligned}$$

I noen tilfeller med digrafer vil det kunne oppstå konflikt mellom ulike regler:

$$\begin{aligned} \langle \text{egne} \rangle &\rightarrow \{\text{egne}\} \mid \langle \text{e} \rangle \rightarrow \{\text{e}\} + \langle \text{g} \rangle \rightarrow \{\text{g}\} + \langle \text{n} \rangle \rightarrow \{\text{n}\} + \langle \text{e} \rangle \rightarrow \{\text{e}\} \\ \{\text{egne}\} &\rightarrow /e:gne/ \mid \{\text{VK}_1+\text{K}_2\} \rightarrow [\text{V}:\text{K}_1\text{K}_2] \end{aligned}$$

$$\langle \text{egne} \rangle \rightarrow \{\text{æjne}\} \mid \langle \text{eg} \rangle \rightarrow \{\text{æj}\} + \langle \text{n} \rangle \rightarrow \{\text{n}\} + \langle \text{e} \rangle \rightarrow \{\text{e}\}$$

$$\begin{aligned} \langle \text{egne} \rangle &\rightarrow \{\text{e}ñne\} \mid \langle \text{e} \rangle \rightarrow \{\text{e}\} + \langle \text{g} \rangle \mid \text{n} \rightarrow \{\text{ñ}\} + \langle \text{n} \rangle \rightarrow \{\text{n}\} + \langle \text{e} \rangle \rightarrow \{\text{e}\} \\ \{\text{e}ñne\} &\rightarrow /eñne/ \mid \{\text{VK}_1(\neg+)\text{K}_2\} \rightarrow [\text{V}:\text{K}_1\text{K}_2] \end{aligned}$$

Her er de generative reglene flertydige ettersom de ikke har prioritet i forhold til hverandre, og de genererer tre gyldige segmentale fonemsekvenser. Eksemplet viser også at vi med dette regelsettet ville få en overgenerering av fonologiske uttrykk for ordformer som <legning>.

Videre går det fram at reglene må gjøre bruk av morfologisk informasjon for å tilegne korrekt kvantitet. Slik informasjon er selvfølgelig ikke umiddelbart tilgjengelig for leseren.

I dette neste eksemplet er det nødvendig med et regelsett som gjør bruk av flere prioriterte regler der resultatet av den foregående blir utgangspunktet for den neste.

1. $\langle \text{godt} \rangle \rightarrow \{\text{gott}\} \mid \langle \text{g} \rangle \mid \text{V}_{\text{bakre}} \rightarrow \{\text{g}\} + \langle \text{o} \rangle \rightarrow \{\text{o}\} + \langle \text{dt} \rangle \# \rightarrow \{\text{tt}\} \#$
2. $\{\text{gott}\} \rightarrow \{\text{gått}\} \mid \{\text{o}\} \mid \text{tt} \# \rightarrow \{\text{å}\}$
3. $\{\text{gått}\} \rightarrow /gått/ \mid \text{V} \mid \text{KK} \# \rightarrow \text{V}:\text{KK}$

For å generere riktige uttaler for <gode>_{a=}^{/2}go:e/ og <gode>_{s=}^{/2}go:de/, trenger reglene å bruke informasjon om bøyningsjunktur og morfemjunktur.

Til slutt et eksempel som gjør bruk av leksikalsk informasjon:

<seksten>	→ {seksten}	<s> → /s/ + ...
{seks}	→ /sæjs/	_{ten} ∧ {seksten} = "16"
{ten}	→ /tɲ/	trykksvak _
	→ /sæjstɲ/	

Det kan synes som om regelen som genererer diftongen, er unødvendig spesifikk, og at det ville vært tilstrekkelig med en fonologisk kontekst, eller i det minste en junkturkontekst. For dagens SØ leksikon ville en kontekstbasert regel være tilstrekkelig. Det er imidlertid åpenbart at dette ikke er en produkt regel, men en regel som hefter ved akkurat dette leksemet eller denne stammen. Om det mot formodning skulle dukke opp en homograf til dette ordet, for eksempel en ny høyteknologisk rokk med seks tener, altså en *seksten*, bør regelen ikke trigges av et slikt nyord. Dette sikrer vi ved å legge inn et leksikalsk kriterium.

7 Kvantitative undersøkelser

7.1 Grunnleggende diskusjoner

Til slutt vil jeg presentere og diskutere noen kvantitative undersøkelser jeg har gjort på grunnlag av ord fra *Norsk Uttaleordbok* og ord fra frekvensordlistene fra Oslo-korpuset. Så langt har jeg i hovedsak behandlet bokmålet som et system uten å ta mye hensyn til hva som er sentrale egenskaper og hva som er perifere egenskaper, hva som er frekvente mekanismer og hva som er sjeldne mekanismer. Jeg har riktignok referert kort til små stikkprøver rundt hvilke ord som er i bruk, og hvilke ord som blir feilstavet, men ellers har jeg diskutert systemet.

Jeg har brukt ord som system, struktur, mønster og mekanismer uten å referere til frekvens, og jeg har brukt ordet regelmessighet, som jeg har definert i relasjon til struktur. Ordet *regulær*, eller snarere *irregulær*, har jeg bare brukt i en gjengivelse av et arbeid av Danielsson 2003 fordi dette var det ordet hun brukte om grafonomiske tegn av typen <og>=/å:/ og <seksten>=/²sæjsten/.¹ Men ordet *regularitet* har jeg ikke definert.

Jeg var også inne på denne diskusjonen i definisjonen av grafemet og primærfonem, om frekvens kunne brukes til å definere regularitet som for eksempel i relasjonen mellom grafem og fonem i det grafemiske tegnet. Jeg ønsket derfor å gjøre noen enkle kvantitative undersøkelser for å se om det vi oppfatter som "regler" og "unntak" bare forholder seg til språkstrukturen eller om de har noen relasjon til frekvens.

Jeg har derfor valgt noen problemstillinger som er diskutert tidligere i oppgaven til dette formål:

1. Vokalfonemer og grafemer: <u>:<o>, <o>:<å>, <e>:<æ>
2. Diftonger og digrafer: <ei>:<eg>

¹Dette var mine egne eksempler fra norsk. Danielsson skrev om svensk.

3. Prioritering av digrafer: <gn>
4. Fonemer uten primærgrafem: /ʃ/
5. Kvantitet og dobbeltkonsonant.

7.2 Data

7.2.1 Om bruk av korpus

Jeg har brukt bokmålsdelen av det såkalte Oslo-korpuset,¹ som består av 18,5 millioner ord fra aviser, skjønnlitteratur og sakprosa. Til denne undersøkelsen har jeg brukt de frekvensordlistene som er utarbeidet på grunnlag av avisdelen av korpuset, og som består av bare 9,5 millioner ord.

Selv om dette er en relativt stor samling av ord, og fra tre ulike typer tekster, må man ikke se bort fra det faktum at det representerer et svært snevert utvalg av språkbruk, noe som kan illustreres om man studerer listen over de mest frekvente substantiver. De tre mest brukte substantivleksemer i avisdelen av dette korpuset er *år*, *krone* og *dag*, fulgt av *tid* og *bergen*. Jeg tror ikke det skal være nødvendig å utføre omfattende undersøkelser for å underbygge en påstand om at denne fordelingen er temmelig ulik fordelingen i en språkbrukers dagligspråk, i skjønnlitterært språk eller i vitenskapelige avhandlinger.

Et annet interessant eksempel på statistiske feil som kan stamme fra bruk av frekvensordlister, kommer fram om man sammenligner frekvenslistene med *Nynorsk frekvensordbok* (Vestbøstad 1989), som har hentet data delvis fra nynorske dags- og lokalaviser og delvis fra ukepresse og skjønnlitteratur. Sandøy 2000, som skriver om importord i norsk, rapporterer med henvisning til nettopp *Nynorsk frekvensordbok* at det mest brukte importordet av fransk opprinnelse i norsk, nemlig *kommune*, er vesentlig mer brukt enn det mest brukte importordet av engelsk opprinnelse i norsk, nemlig *starte*. *Kommune* kommer på 77. plass i

¹ <http://www.tekstlab.uio.no/norsk/bokmaal/>

ordlista, mens *starte* kommer på 304. plass. Det er særlig to merknader som bør gjøres her:

1. Kildene for ulike frekvensordlister vil være ulike, og disse ulikhetene vil gjenspeiles i resultatene. Kildene til Vestbøstad 1989 består av 50% tekster fra nynorske dags- og lokalaviser, og populariteten til et ord som *kommune* kan naturligvis skrive seg fra hva lokalaviser på vestlandet er opptatt av.
2. Frekvensordlister kan sorteres etter ordformer eller etter leksem, og dette kan utgjøre store forskjeller. *Nynorsk frekvensordbok* er ordnet etter leksem, mens de lettest tilgjengelige listene under Oslo-korpuset er ordnet etter ordform, og her kan det oppstå store skjevheter i sammenligningen mellom leksem med komplekse bøyningssystemer og leksem med enkle bøyningssystemer. Dessuten kan det i mange sammenhenger, for eksempel i undersøkelsen til Sandøy 2000, være mer interessant å sammenligne morfemenes frekvens, og ikke leksemene. I praksis vil dette si å slå sammen tallene for substantivet *start* og verbet *starte*, i tillegg til sammensetninger som *oppstart* og *startplass* på den ene side, og rota *kommune* og avledninger som *kommunal* og sammensetninger som *kommunesammenslåing* på den andre. Det er ikke gitt at resultatet blir det samme om man gjør det på denne måten; det er store ulikheter mellom ulike morfemers evne eller potensial til å inngå i avledninger og sammensetninger.
3. Sandøy 2000 konkluderer på vegne av norsk ved å vise til en undersøkelse om nynorsk. Det kan selvfølgelig være store ulikheter mellom målformene når det gjelder ordvalg, avhengige blant annet av hvilke synonymer eller nesten-synonymer som er tilgjengelige som konkurrerende valgmuligheter. I dette tilfellet vil bokmålsbrukere gjerne kunne velge å bruke *begynne*¹, mens nynorskbrukere kanskje i større grad

¹ Leksemet *begynne* har 4011 forekomster i avisdelen av Oslo-korpuset, mens *starte* har 3220.

vil velge å bruke *byrje*¹. Begge målformene vil også kunne velge alternative fraser som *sette i gang*, som er vesentlig vanskeligere å spore opp. Mange ulike variabler vil styre disse ulike valgene, og det blir dermed vanskelig å bruke undersøkelser om nynorsk til å konkludere om norsk språkbruk generelt.²

7.2.2 Datautvalg

Jeg har gjort et utvalg av oppslagsord fra *Norsk Uttaleordbok* (NUOV) ved å ta det øverste ordet i hver kolonne, totalt 645 ord. Dette er for såvidt ikke noe statistisk tilfeldig utvalg, men det er en metode som sikrer et visst utvalg av ulike typer ord. Imidlertid har jo ordboksredaktøren gjort et utvalg før meg, og hans metoder og intensjoner kjenner jeg ikke, men utvalget inneholder uten tvil et spekter av ord av ulik bruksfrekvens. Ordboken inneholder kun oppslagsformer, så dette datautvalget gir kun et utvalg av leksemer, eller av ubøyde ordformer.

Jeg har videre to frekvensordlister fra avisdelen av Oslo-korpuset. Avisdelen er på totalt 9 520 502 ord og 376 531 ulike ordformer, og listene er en liste med de 1000 mest frekvente ordene fra korpuset og en liste med de 10,000 mest frekvente ordene. Den korte listen representerer 6,0 millioner ord i løpende tekst, eller omtrent 63%, mens den lange listen representerer 7,8 millioner ord i løpende tekst, omtrent 82% av hele korpuset. Disse listene inneholder ordformer fra løpende tekst, slik at det her er snakk om (også) bøyde former. Det betyr at frekvente verb og substantiver vil være til stede i listene flere ganger med ulike bøyninger; et tall for antall ulike ordformer i et søkeresultat er dermed ikke et tall for antall ulike leksemer. Listene inneholder dessuten en del forkortelser og navn, både person-, bedrifts- og geografiske navn, og noen ord som ikke er bokmålsord, i hovedsak nynorskord. Disse ordene er med i utvalgene, men

¹ Leksemet *byrje* har 197 forekomster i *Nynorsk frekvensordbok*, mens *starte* har 291 og *begynne* 72.

² Det motsatte gjelder selvfølgelig også; man kan heller ikke bruke undersøkelser om bokmål til å konkludere om norsk språkbruk generelt. Imidlertid utgjør bokmålstekster en så mye større andel av den totale mengden norske tekster at en slik eventuell konklusjon i større grad vil kunne sees som gyldig. Min undersøkelse dreier seg spesifikt om bokmål, så for denne undersøkelsen oppstår det ikke slike problemer.

forkortelsene og ikke-bokmålsordene er forsøkt fjernet når det er mulig; i alle fall spiller de ingen stor rolle i tallmaterialet. Navnene er representert i tallmaterialet, med unntak av bedriftsnavn som er forkortelser. Forskjellen mellom versaler og minuskler er fjernet fra listene, slik at i prinsippet er det umulig å vite om <fred> representerer /fre:d/="ro" eller /fred/="mannsnavn".

Frekvensordlistene består av en linje for hver ordform med et tall foran som representerer nummer i listen, og et tall etter som representerer antall forekomster i korpuset.

7.3 Metode

7.3.1 Transskripsjon

Ordbokslisten og den korte frekvenslisten er håndtransskribert av meg, med ett oppslagsord pluss transskripsjon på samme linje. For ordbokslisten har jeg brukt ordbokens transskripsjon, oversatt til et rent ASCII-basert transskripsjonssystem for å lette søkingen. Der ordboken har hatt flere varianter, har jeg ført opp alle i sin helhet etter hverandre på samme linje, også dette for å gjøre søkingen lettest mulig. Der jeg har hatt en annen og begrunnet intuisjon enn ordboken, har jeg også ført opp den, men merket den spesielt, slik at jeg kunne velge å fjerne den automatisk senere. Det er et lite antall ord dette gjelder.

For den korte frekvensordlisten har jeg brukt egen intuisjon, men støttet meg på NUOV når jeg har vært i tvil. Dette betyr at det ikke er sikkert at alle transskripsjoner er i tråd med Vanvik, men jeg har kunnet konstatere at vi i all hovedsak er enige, og jeg har også kunnet oppdage en tendens med hensyn til i hva slags tilfeller vi ikke er enige, slik at jeg har kunnet slå opp disse tilfellene, samt de tilfellene der jeg har vært i tvil. Også her har jeg ført opp alle Vanviks varianter og eventuelt mine egne. Imidlertid er det to store forskjeller i transskripsjonen av denne listen i forhold til listen fra ordboken.

- Frekvensordlistene inneholder bøyde ordformer, som ikke er transskribert i NUOV. Jeg har derfor i mye større grad måttet støtte meg på egen intuisjon, særlig når det gjelder kvantitet og assimilasjoner.
- Frekvensordlistene inneholder ingen informasjon om hvilket leksem hver ordform representerer; i prinsippet vil hver grafisk form i en liste representere alle homografer fra ulike leksemer. I praksis er hvert leksem representert i en andel av hver form som reflekterer hvert leksems frekvens, men jeg har ikke hatt tid til å fordele hver ordform mellom leksemene etter relativ frekvens, så jeg har transskribert hver homograf med alle mulige fonemiske representasjoner og ført dem opp på samme linje, akkurat som variantene i ordbokslisten. Noen ganger er det åpenbart at dette lager skjevheter i materialet, og noen slike skjevheter har jeg justert manuelt ved opptelling.¹

Etter transskripsjon har jeg laget en ekstra kolonne med en segmental transskripsjon der alle symboler for tonem, trykk, kvantitet, stavelsesgrense og merke for stavelsesbærende konsonant er fjernet. Slik har hver linje bestått av først den grafiske formen, dernest den komplette fonemiske transskripsjonen med alle varianter, og til slutt den segmentalfonemiske transskripsjonen med alle varianter, i tillegg til tallene som nevnt over. Disse tre delene har jeg adskilt med spesielle merker, slik at jeg kunne spesifisere hvilken kolonne et søk skulle gjelde.

Jeg har ikke transskribert den lange lista på 10,000 ord, men jeg har brukt den til å søke etter grafemsekvenser i, slik at jeg deretter kunne transskribere søkeresultatet med hensyn til den variabelen jeg skulle undersøke. For eksempel søkte jeg etter ord med <gn> i, og deretter kunne jeg undersøke de 78 ordformene i resultatlista og bestemme hvilke av disse som hadde <gn>=/gn/, hvilke som hadde <gn>=/ɲn/, og hvilke som hadde e<gn>=/æjn/.

¹Ett slikt eksempel er der <fred>, som neppe representerer mannsnavnet *Fred* særlig frekvent.

7.3.2 Søking

Også for dette formålet har jeg søkt i tekstfilene ved hjelp av regulære uttrykk med et grep-program. Her har jeg kunnet spesifisere kriterier i grafemdelen og fonemdelene av filene samtidig og slik få ulike utvalg av grafem-fonem-korrespondanser. Disse to søkene er gjort i ordbokslista:

```
. *rs.*;2.*3;.*S.*\r\n
```

```
. *. *;2;.*[12]([bcdfghjklmnpqrstvwxyz]){0,3}o[^:]. *;3;.*.*;.*\r\n
```

Det første søker etter ordformer med <rs>=/ʃ/, mens det andre søker etter kort /o/ i trykksterk stavelse. Ved å skille kolonnene med full transskripsjon og segmentaltransskripsjon, kan jeg velge å søke i den siste kolonnen når det bare er segmentale enheter jeg er interessert i, og slik unngå å måtte ta hensyn til tonem- og lengdemerker som kan opptre inni fonemsekvenser.

I et søk som det første, med korrespondanse mellom grafemer og fonemer, inneholder gjerne resultatet også noen "søppeltreff", altså treff som tilfredsstiller kriteriene, men ikke intensjonen bak kriteriene. I dette eksemplet kunne det være ord der <rs> ikke representerte /ʃ/, men som inneholdt en <sj>-digraf i tillegg. Dersom trefflista ikke er for lang, er det mest økonomisk å prøve å fjerne søppeltreff manuelt etter søket framfor å gjøre søkekriteriene mer presise.

Det er to viktige kilder til unøyaktigheter her:

- Dersom en søkt egenskap forekommer to ganger i et ord, blir den likevel telt bare én gang. <nærvær>=[nærvær] gir altså som resultat bare én <æ>=/æ/-korrespondanse, med mindre søket spesifiserer at det skal være to.
- Dersom en søkt egenskap forekommer i én uttalevariant av en grafisk ordform, men ikke en annen, også dersom variantene egentlig skriver seg fra homografi, blir denne ordformen også telt én gang, selv om kanskje den trufne varianten er den minst frekvente.

Det siste punktet kan føre til at en ordform blir telt på begge sider i en sammenligning; for eksempel vil <egne> bli telt som <gn>=/gn/ og som <gn>=/ɲn/, og begge to vil bidra like mye i resultatet.

I ordbokslista er ikke denne problemstillingen så aktuell, i og med at her står hver ordform for bare ett leksem.

Det er også en litt mer prinsipiell problemstilling rundt telling, og det er ved tvetydigheter i korrespondansene: Lisensierer for eksempel /ʃ/ <rs> eller <s>|_l i <forslag>=/²fåʃlag/? Jeg har valgt en pragmatisk løsning på dette og telt den med en gang for hver, akkurat som for uttalevariantene.

7.3.3 Opptelling og beregning

Etter søking er resultatene hentet inn i regnearkprogrammet Calc for behandling. Resultatene fra ordbokslista innebærer en ren leksemtelling og er ikke så komplisert.

Resultatene fra frekvenslistene innebærer imidlertid muligheter for å ta hver ordforms frekvens med i beregningen, i tillegg til å behandle dem som ordlistedata. Jeg har telt både ordformstyper og -eksemplarer og foretatt enkle prosentberegninger av resultatene.

7.4 Resultater

7.4.1 Ordbokssøk

	/å:/	/å./	/å/	/å/
<å>	18	5	23	20%
<o>	11	81	92	80%
Sum	29	86	125	

	<i>/o:/</i>	<i>/o./</i>	<i>/o/</i>	<i>/o/</i>
<o>	35	8	43	93%
<u>	0	3	3	7%
Sum	35	11	46	
	<i>/æ:/</i>	<i>/æ./</i>	<i>/æ/</i>	<i>/æ/</i>
<æ>	7	0	7	23%
<e>	1	22	23	77%
Sum	8	22	30	

	<i>/æj/</i>	
<ei>	15	83%
<eg>	2	11%
<ay>	1	6%
Sum	18	

	<i><gn></i>	
/gn/	4	50%
/æjn/	2	25%
/ɨn/	1	13%
/nj/	1	13%
Sum	8	

	<i>/ʃ/</i>	
<rs>	19	32%
<sj>	17	28%
<s>_l	10	17%
<skj>	5	8%
<sk>	3	5%
<g>	2	3%
<ch>	2	3%
<c>	1	2%
<s>	1	2%
Sum	60	

Antall treff i denne lista er så lite at det har vært enkelt å justere resultatene manuelt i etterkant av søkene. Antall treff er også så lite at det er problematisk å trekke konklusjoner ut fra så mangelfulle statistiske data. Jeg vil likevel trekke fram noen resultater:

Bare 6% av korte /å/-fonemer lisensierer <å>; resten lisensierer <o>.

Bare 32% av <o>-grafemene brukes til å representere /o/; resten representerer /å/.

Alle lange /o/-fonemer lisensierer <o>.

Ingen korte /æ/-fonemer lisensierer <æ>.

Blant bare 60 leksemer med /ʃ/ fant jeg 9 ulike representasjoner.¹

7.4.2 Frekvensordlister

Vokaler / diftonger og digrafer

	<e>	<æ>	Sum
/e/	2242416	0	2242416
/æ/	275563	44191	319754
Sum	2517979	44191	2562170
14% av /æ/ lisensierer <æ>			
89% av <e> representerer /e/			

	<u>	<o>	Sum
/u/	213361		213361
/o/	4857	287419	292276
Sum	218218	287419	505637
98% av /o/ lisensierer <o>			
98% av <u> representerer /u/			

	<o>	<å>	Sum
/o/	287419		287419
/å/	887783	565808	1453591
Sum	1175202	565808	1741010
39% av /å/ lisensierer <å>			
24% av <o> representerer /o/			

	<eg>	<ei>	Sum
/eg/	18444	0	18444
/æj/	103784	10532	114316
Sum	122228	10532	132760
9% av /æj/ lisensierer <ei>			
15% av <eg> representerer <eg>			

¹Da er riktignok <s> telt to ganger avhengig av grafonomisk kontekst.

Tallene om sammenheng mellom vokalfonemer og grafemer, samt om diftongen /æj/, er gjort ved å søke etter kombinasjonene av disse to variablene i den korte frekvensordlista. Her er det ikke skilt mellom lange og korte vokaler, og det er bare telt eksemplarer og ikke typer. Vi kan likevel se at tallene bekrefter noen av tendensene fra ordbokssøket. De viktigste prosentandelene er gitt i tabellene, og om vi utvider *markør*-begrepet litt for dette formålet, kan vi si at:

- <e> er en ganske sterk (89%) markør for /e/, mens <æ> ikke blant de 1000 mest frekvente ordene representerer /e/ i det hele tatt, selv om vi i de fonemografiske oversiktene har gitt mange eksempler på nettopp dette. <æ> er altså en sterk (100%), men svært partiell (14%), markør for /æ/.
- <u>=/u/ er en svært regulær korrespondanse, med styrke på 98% og totalitet (altså 100%).
- <o> er en svært svak (24%) markør for /o/, og <å> er en veldig partiell (39%) markør for /å/. Om vi hadde begrenset søket til bare korte /å/er, hadde resultatet trolig blitt enda mer slående.
- <eg> er både en sterk (85%) og en nesten total (89%) markør for /æj/. På den annen side er <ei>, som vi kunne vente, en fullkomment sterk markør for /æj/.

Vi ser at for mappingene /æ/ → <æ>, <o> → /o/, /å/ → <å> og /æj/ → <ei> strider resultatene mot de antagelsene jeg gjorde om primærgrafem og primærfonem. Om disse begrepene skal defineres på annen måte enn ved opplisting, må definisjonen baseres på andre egenskaper enn frekvens.

Prioritering mellom digrafer

<gn> i 1-10000

	typer	eksemplarer	typer i %	eksemplarer i %
/gn/	32	7688	41	40 %
/ŋn/	29	6669	37	35 %
/jn/	12	3602	15	19 %
/n/	3	824	4	4 %
annet	2	533	3	3 %
SUM	78	19316	100	100

Tallene for <gn> er kommet fram gjennom søk på grafemsekvensen i den lange frekvensordlista, for så i etterkant å bestemme hvilken kategori av

fonemkorrespondanse hver ordform tilhørte. /jn/ i tabellen representerer sekvensen diftong+/n/, som i <regn>=/ræjn/ og <døgn>=/døjn/, og tabellen gir resultater telt både etter typer og etter eksemplarer i korpuset. Det går tydelig fram av tallene at mappingene til /gn/, /ɲn/ og /Dn/ alle må regnes som regulære. Jeg har ikke undersøkt relasjonene i motsatt retning.

Fonemer uten primærgrafem

/ʃ/ i 1-1000

<>	typer	eksemplarer	typer i %	eksemplarer i %
rs	27	67245	46	50 %
sl	15	36912	25	27 %
skj	6	12412	10	9 %
sj	7	8136	12	6 %
s	2	7795	3	6 %
sk	2	1901	3	1 %
SUM	59	134401	100	100

I diskusjonen om primærgrafem på s.77 slo jeg fast at ikke alle fonemer har noe entydig primærgrafem. For å undersøke situasjonen for /ʃ/ har jeg søkt etter dette fonemet i den korte frekvensordlista og deretter funnet det korresponderende grafemiske uttrykket. Denne lyden kalles gjerne "skj-lyden" i pedagogisk sammenheng, noe som stemmer dårlig med det faktum at bare 10% av typene og enda færre av eksemplarene lisensierer denne trigrafen.

Imidlertid er det, som nevnt, mulig å bruke et dypere fonembegrep og klassifisere [ʃ]_1 som en allofon av /s/ og <rs>=[ʃ] som et assimilasjonsprodukt av /rs/. I så fall er <skj> den mest totale markøren for "skj-lyden", i hvert fall på eksemplarnivå. På typenivå er <sj> den mest totale markøren. Dette er imidlertid ikke i tråd med den fonemanalysen jeg har lagt til grunn, og heller ikke i tråd med Vanvik (1979 og 1985). Uansett er <skj> utenom morfemgrense en fullstendig sterk markør for /ʃ/; den representerer aldri andre fonemer.

Kvantitet og dobbeltkonsonant

		80 %		
Eksemplar	Lang vokal	Kort vokal	Sum	
55 %	<K>	828575	1000498	1829073
	<K ₁ K ₂ >	27420	81263	108683
	<K ₁ K ₁ >	0	174702	174702
	Sum	855995	1256463	2112458

		34 %		
Type	Lang vokal	Kort vokal	Sum	
32 %	<K>	73	34	107
	<K ₁ K ₂ >	12	23	35
	<K ₁ K ₁ >	0	43	43
	Sum	85	100	185

I undersøkelsen av kvantitet har jeg søkt etter korrespondanser mellom kort og lang vokal og enkeltkonsonant (K), konsonantsekvens (K₁K₂) og reduplisert konsonant (K₁K₁) i enstavelsesord. Summen av konsonantsekvens og reduplisert konsonant kaller jeg dobbeltkonsonant.

Kanskje noe overraskende fant jeg at på eksemplarnivå er enkeltkonsonant en *sterkere* markør (55%) for kort vokal enn for lang vokal. Også overraskende var at enkeltkonsonant er en *vesentlig mer total* markør (80%) for kort vokal enn dobbeltkonsonant. I allmennspråket betyr det at de fleste enkeltkonsonanter står etter korte vokaler, og at de fleste korte vokaler i enstavelsesord i løpende tekst ikke er markert!

Det er mange forbehold å ta her i tillegg til de generelle om ordutvalget i listene. Søket etter bare enstavelsesord resulterer i at mange bøyde former av verb og substantiv som har den regelmessige korrespondansen mellom mellom kvantitet og dobbeltkonsonant, ikke blir med i søket. Akkurat i dette søket gjør nok det smale ordutvalget større utslag enn i de fleste andre søkene over, kanskje med unntak av <eg>=/æj/-korrespondansen. Kort vokal etterfulgt av enkeltkonsonant er et trekk som er typisk for frekvente "småord". Søket inkluderer ord med kortvokal og <m>#, som ifølge den grafotaktiske begrensningen ¬<mm># er regelmessig.¹

¹Men kort vokal etterfulgt av <m># bidrar uansett til det visuelle mønsteret vi tilegner oss som lesere, og som danner grunnlaget for lese- og skriveferdighetene hos barn.

7.5 Diskusjon

Tallene fra ordboka bygger på bare 645 leksemer, og det er begrenset hvilke konklusjoner man kan trekke ut fra det. De små tallene gir store utslag av statistisk tilfeldighet. Imidlertid ser vi også spor av at ordlisteundersøkelsene gir et mer representativt utvalg siden også mindre frekvente ord er med.

Jeg synes likevel resultatene fra frekvenslistene er mer interessante; disse representerer ordformer man vet er i bruk – i hvert fall i avisene. Alle de frekvente funksjonsordene vil være til stede, noe de ikke er i ordbokslista, og disse ordene er representert i alle typer tekster språkbrukere møter som lesere og skrivere. Ekstra interessante er kanskje de søkene som er basert på den lange listen. De 10,000 mest frekvente ord representerer et viktig utdrag av løpende tekst, selv om Oslo-korpuset også inneholder 366,000 ordformer som ikke er med i undersøkelsen.

Likevel er det viktig å være oppmerksom på at listene representerer et utvalg. I frekvensordlistene finner vi alle de høyfrekvente ordene som vi *vet* har uregelmessige stavemåter, og den statistiske fordelingen mellom regelmessige og uregelmessige stavemåter vil være annerledes i resten av teksten. Jeg har også fokusert på telling av eksemplarer og ikke typer. Hvis formålet med oppgaven er å undersøke språkbrukerens intuisjon om mønstrene i normen, er kanskje telling av eksemplarer like interessant. Vi vet lite om hvilken rolle størrelser på typenivå og størrelser på eksemplarnivå spiller i vår kognitive oppfattelse av mønster i et (skrift-)språkssystem. Imidlertid har sannsynligvis de mest frekvente ordformene den viktigste rollen i danning av mønster i den grunnleggende lese- og skriveopplæringen, og ikke de hundretusener av andre ordformer.

Denne oppgaven er en undersøkelse av bokmålsnormen og ikke et innlegg i normeringsdebatten. Som nevnt har normeringsinstansene også andre variabler å ta hensyn til, i tillegg til dem jeg har diskutert i denne oppgaven; særlig dialektvariasjon og tradisjon spiller viktige roller her. Jeg undrer meg likevel over at et så uregelmessig trekk som enkeltkonsonantene etter kort vokal i frekvente

småord ble værende i normen i en tid da det hersket en såpass stor vilje til reform og det likevel ble foretatt såpass store omveltninger i normen.

Feil bruk av dobbeltkonsonant er et av de mest omfattende problemområdene for barn, og det er også, som vi har sett av noen stikkprøver i Usenet-korpuset, et problemområde for mange voksne. *<værre> er det mest frekvente tilfellet av feilstaving jeg har funnet i Usenet-korpuset. Dersom store deler av ordformene i de tekstene vi møter både i oppveksten og i voksen alder, slett ikke forholder seg til "reglene" i disse to områdene av rettskrivningen, er kanskje ikke det så rart. At de frekvente uregelmessige ordformene beherskes av de fleste, er naturlig, ettersom frekvensen fører til at de danner et kjent visuelt inntrykk hos språkbrukeren. Men det er sannsynligvis slik at de etterlater spor som dermed fører til vanskeligheter for mange språkbrukere med å bruke de kodifiserte normene for mindre frekvente ordformer.¹ Jeg tror det er viktig å rette oppmerksomhet mot dette i den grunnleggende lese- og skriveopplæringen.

¹Også "feilstavede" former som *<værre> etterlater sannsynligvis slike spor.

Litteratur

- Allén, Sture 1971: Introduktion i grafonomi. Almqvist & Wiksell.
- Austad, Ingolv (red.) 1997a: Mening i tekst. Cappelen.
- Austad, Ingolv 1997b: Lesing som forståelse. I: Austad 1997a.
- Aars, J. 1907: Ordlister til norsk retskrivning. Fabritius.
- Berulfsen, Bjarne 1969: Norsk Uttaleordbok. Aschehoug.
- Blaauw, Knud (red.) 1993: Standardspråk og dialekt – Seminarer i Oslo 1991 og 1992. Bergens riksmålsforening / Det norske akademi for sprog og litteratur.
- Borgstrøm, Carl Hj. 1958: Innføring i sprogvidenskap. Universitetsforlaget.
- Brodersen, Randi Benedikte; Kinn, Torodd (red.) 2000: Språkvitskap og vitskapsteori - ti nye vitskapsteoretiske innlegg. Ariadne.
- Brunstad, Endre 2000: Språknormer som forskningsobjekt. I: Brodersen og Kinn 2000.
- Bråten, Ivar 1990: Kognitive strategier og ortografi. Pedagogisk forskningsinstitutt, Universitetet i Oslo.
- Bråten, Ivar 1994: Skriftspråkets psykologi. Høyskoleforlaget.
- Chomsky, Noam; Halle, Morris 1968: The sound pattern of English. MIT Press.
- Coulmas, Florian 2003: Writing Systems. Cambridge University Press.
- Danielsson, Kristina 2003: The relationship between grapheme-phoneme correspo. I: Scandinavian journal of educational research.
- DeFrancis, John 1989: Visible Speech. University of Hawaii Press.
- Durand, Jacques 1990: Generative and non-linear phonology. Longman.

- Dysthe, Olga 1993: Ord på nye spor. Samlaget.
- Dyvik, Helge 1993: Normbegreper, systembegreper og individuering av norske skriftspråk. I: Blaauw 1993.
- Dyvik, Helge 2003: Offisiell og ikke-offisiell språknormering – nyttig eller skadelig motsetning? I: Omdal og Røstad 2003.
- Eitrem, H.; Sandvei, M. 1960: Rettskrivning etter ny læreboknormal 1959. Fabritius.
- Elvemo, Jarle 1986: Leseprosessen - vansker og metoder, en kognitiv psykologisk innføring. Gyldendal.
- Faarlund, Jan Terje; Lie, Svein; Vannebo, Kjell Ivar 1997: Norsk referansegrammatikk. Universitetsforlaget.
- Gelb, I.J. 1963: A study of writing. University of Chicago Press.
- Hagtvet, Bente Eriksen 1996: Skrivelyst og språklig bevissthet. I: Wold 1996.
- Harris, Roy 1995: Signs of writing. Routledge.
- Haslev, Marianne 1972: Morfo-fonemikk. Universitetsforlaget.
- Haslev, Marianne 1985: LORANg eller den forbudte fonologi. Bind 1. Institutt for fonetikk og lingvistikk, Universitetet i Bergen.
- Hellmark, Christian 2000: Typografisk håndbok. Spartacus.
- Høien, Torleiv; Lundberg, Ingvar 1991: Dysleksi. Gyldendal.
- Jakobson, Roman; Halle, Morris 1956: Fundamentals of language. Mouton.
- Jones, Daniel 1991: English Pronouncing Dictionary. Cambridge University Press.
- Katamba, Francis 1989: An introduction to phonology. Longman.

- Katzner, Kenneth 1986: The languages of the world. Routledge.
- Knudsen, Trygve; Sommerfelt, Alf 1983: Norsk Riksmålsordbok. Kunnskapsforlaget.
- Kristoffersen, Gjert 2000: The phonology of Norwegian. Oxford University Press.
- Krogsrud, Torgeir; Seip, Didrik Arup 1942: Norsk Rettskrivningslære - etter rettskrivningen av 1938. Steenske forlag.
- Landrø, Marit Ingeborg; Wangensteen, Boye (red.) 1993: Bokmålsordboka. Universitetsforlaget.
- Lundeby, Einar 1987: Fremmedord - hva gjør vi med dem?. I: Mæhle, Lundeby og Grønvik 1987.
- Mæhle, Leif; Lundeby, Einar; Grønvik, Oddrun (red.) 1987: Fornying og tradisjon - Språkvern og språkrøkt 1972-1988. Norsk Språkråd / Cappelen.
- Nguyễn Đình-Hoà 1991: Vietnamese-English dictionary. Tuttle.
- Omdal, Helge; Røstad, Rune (red.) 2003: Krefter og motkrefter i språknormeringa - om språknormer i teori og praksis. Høyskoleforlaget.
- Oskal, Inger Marie 2002: Læreplanordbok - norsk-samisk ordliste til L97S. Sámedikki oahpahuossodat.
- Papazian, Eric 1984: Talemål og skriftspråk. Institutt for fonetikk og lingvistikk, Universitetet i Bergen.
- Partee, Barbara H.; ter Meulen, Alice; Wall, Robert E. 1993: Mathematical methods in linguistics. Kluwer.
- Sampson, Geoffrey 1985: Writing Systems. Stanford University Press.

- Sandøy, Helge 2000: Lånte fjører eller bunad? - om importord i norsk. LNU / Cappelen.
- Saussure, Ferdinand de 1959: Course in general linguistics. McGraw-Hill.
- Sivertsen, Eva 1967: Fonologi - fonetikk og fonemikk for språkstudenter. Universitetsforlaget.
- Sproat, Richard 2000: A computational theory of writing systems. Cambridge University Press.
- Svennevig, J.; Sandvik, M.; Vagle, W. 1995: Tilnærminger til tekst - modeller for språklig tekstanalyse. LNU / Cappelen.
- Torp, Arne; Vikør, Lars 2003: Hovuddrag i norsk språkhistorie. Gyldendal.
- Vanvik, Arne 1979: Norsk fonetikk. Fonetisk institutt, Universitetet i Oslo.
- Vanvik, Arne 1985: Norsk Uttaleordbok. Fonetisk institutt, Universitetet i Oslo.
- Venezky, Richard L. 1970: The Structure of English Orthography. Mouton.
- Vestbøstad, Per (red.) 1989: Nynorsk frekvensordbok. Alma Mater.
- Vikør, Lars 1994: Språkplanlegging - prinsipp og praksis. Novus.
- Vinje, Finn-Erik 2004: Skriveregler. Aschehoug.
- Wiggen, Geirr 1992: Rettskrivningsstudier II. Universitetet i Oslo.
- Wold, Astri Heen (red.) 1996: Skriftspråkutvikling - Om hvordan barn lærer å lese og skrive. Cappelen.

Importord

Listen inneholder etymologiske opplysninger om importord som er omtalt i oppgaven.

allé	fra fransk <i>aller</i> .
avis	fra fransk av latin <i>ad+visum</i> .
bajas	gjennom tysk fra italiensk <i>baiazza</i> (NROB).
bayer	fra tysk Bayern. NROB har <i>baier</i> .
bil	forkortelse av <i>automobil</i> fra gresk <i>autos</i> og latin <i>mobilis</i> .
celle	fra latin.
design	gjennom engelsk fra fransk, opphavelig latin <i>de+signum</i> .
double	engelsk.
fans	av engelsk <i>fanatic</i> .
farse	gjennom fransk fra latin.
flott	="flytende", fra nederlandsk <i>vlot</i> ..
gøy	fra nederlandsk <i>guich</i> , gjennom engelsk <i>guy</i> .
hai	fra nederlandsk <i>haai</i> .
hangar	fra fransk.
illegal	fra middelalderlatin <i>in+lex/legis</i> .
import	gjennom tysk eller engelsk fra latin <i>in+portare</i> .
irregulær	fra latin <i>in+regula</i> .
jingle	engelsk.
joik	fra samisk <i>juoivat</i> (NROB).
jojo	fra et filippinsk språk, gjennom engelsk <i>yo-yo</i> .
joystick	engelsk.
kai	fra nederlandsk <i>kaai</i> .
kajakk	fra eskimoisk.
kick	engelsk.
know-how	engelsk.
koie	fra nedertysk <i>kôje</i> (NBOB).
køye	fra nederlandsk <i>kooi</i> .

lavvo	fra samisk <i>lávvu</i> (Oskal 2002).
lojal	fra fransk av latin <i>lex</i> .
pizza	italiensk.
provençalsk	fra fransk av <i>Provence</i> .
psyke	fra gresk.
qua	latin.
samurai	japansk.
señor	spansk.
sex	engelsk fra latin av <i>sexus</i> .
skeis	kanskje fra østfrisisk <i>schei</i> eller tysk <i>Scheiss</i> .
skøyter	fra nederlandsk <i>schooien</i> (=skøye).
skøyte	fra nederlandsk <i>schuit</i> gjennom engelsk <i>skate</i> .
soya	fra japansk.
sumo	japansk. I norsk kanskje bare i sammensetninger.
tape	engelsk.
taxi	engelsk av <i>taxicab</i> .
troika	fra russisk тройка= <i>trojka</i> .
watt	etter skotsk familienavn.
xenofob	av gresk <i>xenos</i> + <i>phobia</i> av <i>phobos</i> .
xylofon	av gresk <i>xylon</i> + <i>phone</i> .
yoga	sanskrit.
zen	japansk og kinesisk <i>chann</i> av sanskrit <i>dhyana</i> .

Der ikke annet er oppgitt, er opplysningene fra BMOB.

Register

akronym.....	71
alfabet.....	72, 74
alfabetskrift.....	67
allofoni.....	120
allograf.....	56
allografi.....	60
allomorfer.....	36
apostrof.....	82, 124
arkifonem.....	124
arveord.....	32
ASCII.....	69
aspirasjon.....	43
avlyd.....	45
basisstavemåte.....	77
basisuttale.....	74
bindestrek.....	82, 96
bitrykk.....	40, 43
Bokmålsordboka.....	11
bokstav.....	72, 74
bokstavnavn.....	71, 74
cherokee.....	64
dannet talemål.....	23
dekningsgrad.....	94
diakritika.....	79, 87, 94, 139
diftong.....	120
digraf.....	78, 105
digrafer.....	105
distinktive grafiske trekk.....	62
dobbeltkonsonant.....	166
dybde.....	97
dyp grafonomi.....	98
dypgrafemisk.....	88
dypstruktur.....	97
egentlig skrift.....	55
engelsk.....	27
etymologi.....	71
faktisk norm.....	27
figur.....	56
flapp.....	24, 42
fonemisk transskripsjon.....	9
fonemografi.....	49, 66p.
fonografi.....	65p.
fonotaks.....	44, 114

foreskreven norm.....	27
frekvensordlister.....	18
fremmedord.....	32
funksjonell enhet.....	109
G/F-kontekst.....	91, 101
generativ fonologi.....	42
gjennomsiktighet.....	113
graf.....	56p.
grafem.....	55pp.
grafemikk.....	55
grafemisk tegn.....	56, 58
grafiske trekk.....	69
grafonomi.....	56
grafonomisk tegn.....	51, 55
grafotaks.....	33, 111
grep.....	12
gresk.....	64
hjemlige ord.....	33
homofoner.....	52, 101
homografi.....	39, 102
homonymi.....	101
håndskrift.....	58, 87
ikonisk graf.....	68
importord.....	32
intonasjonsgruppe.....	138
Jakobson og Halle.....	56
japansk.....	66
kana.....	66
kerning.....	87
kinesisk.....	60, 66, 104
kodifisert norm.....	27
konsonantreduplikasjon.....	94p., 104
koreansk.....	68
kvantitet.....	43
Leibniz.....	66
lese- og skrivepsykologi.....	37
lesekoden.....	51
leseretningen.....	88, 93, 101, 114
ligatur.....	87
linearitet.....	84
linearitetsprinsippet.....	68, 72, 105p.
lisensiere.....	67
logografi.....	65
lydskrift.....	52
lånord.....	32
markør.....	89
matematisk notasjon.....	53, 72

mellomrom.....	83
metaskrift.....	69, 80
minuskel.....	62, 73p.
moderate former.....	36
morfem-syllabisk.....	66
morfemografi.....	66, 79
morfofonemiske vekslinger.....	97
morse.....	69
multifunksjonell markør.....	92
musikknotasjon.....	53
normering.....	49
norrønt.....	25
nullmorfem.....	67, 96
nullrepresentasjon.....	148
numerisk grafem.....	79
omlyd.....	45
operativ norm.....	27
Oslo-korpuset.....	11, 16
otherness.....	56
overflatestruktur.....	97
Papazian.....	49
partiell markør.....	92
PC-tastatur.....	81
persepsjonsretningen.....	88
polygraf.....	78, 105
primærfonem.....	67p., 74
primærgrafem.....	68, 77
primært fonologisk innhold.....	67
primært grafonomisk uttrykk.....	68, 77
produksjonsretningen.....	88
prosodemer.....	43
radikale former.....	36
reduplikasjon.....	79
reduplisert konsonant.....	166
regelmessig.....	115, 133
regulære uttrykk.....	11p.
representere.....	67
retrofleks lateral.....	42
russisk.....	45, 94, 97
sandhi.....	45
Saussure.....	48
segmentalhomofoner.....	101, 103
segmentalitet.....	87
semafor.....	69
semasiografi.....	53, 64, 72
skopus.....	89
skriftegn.....	56

skript.....	55
skriveretningen.....	88
SMS.....	74
spansk.....	94
språktegn.....	51, 55
standard østnorsk.....	20, 23
stavelse.....	43, 134
stavelsesbærende konsonant.....	43
stavelsesgrense.....	43
sterk form.....	124
strukturalisme.....	48
svak form.....	124
svak markør.....	92
syllabografi.....	66, 141
syntaktisk grafem.....	79
SØ.....	23
tagging.....	17
tallgrafem.....	71
tallord.....	139
tegnspråk.....	52, 69
tonem.....	43, 93
transskripsjonsklammer.....	7
trekkbasert skrift.....	68
trigraf.....	78, 105
trykk.....	43
ugjennomsiktighet.....	113
urban østnorsk.....	23
Usenet-korpuset.....	11, 16, 31
Usenett.....	74
versal.....	62, 73p.
vietnamesisk.....	78
x-høyde.....	69